

## CRYSTALLIZATION OF ISPA

### FIELD OF THE INVENTION

[001] The present invention relates to a mevalonate pathway enzyme responsible for the synthesis of farnesyl pyrophosphate (FPP), and more specifically to IspA also known as farnesyl pyrophosphate synthase (FPPS) and farnesyl diphosphate synthase (FDPS), referred to herein as IspA. Provided is IspA in crystalline form, methods of forming crystals comprising IspA, methods of using crystals comprising IspA, a crystal structure of IspA, and methods of using the crystal structure.

### BACKGROUND OF THE INVENTION

[002] A general approach to designing inhibitors that are selective for a given protein is to determine how a putative inhibitor interacts with a three dimensional structure of that protein. For this reason it is useful to obtain the protein in crystalline form and perform X-ray diffraction techniques to determine the protein's three-dimensional structure coordinates. Various methods for preparing crystalline proteins are known in the art.

[003] Once protein crystals are produced, crystallographic data can be generated using the crystals to provide useful structural information that assists in the design of small molecules that bind to the active site of the protein and inhibit the protein's activity *in vivo*. If the protein is crystallized as a complex with a ligand, one can determine both the shape of the protein's binding pocket when bound to the ligand, as well as the amino acid residues that are capable of close contact with the ligand. By knowing the shape and amino acid residues comprised in the binding pocket, one may design new ligands that will interact favorably with the protein. With such structural information, available computational methods may be used to predict how strong the ligand binding interaction will be. Such methods aid in the design of inhibitors that bind strongly, as well as selectively to the protein.

### SUMMARY OF THE INVENTION

[004] The present invention is directed to crystals comprising IspA and particularly crystals comprising IspA that have sufficient size and quality to obtain useful information about the structural properties of IspA and molecules or complexes that may associate with IspA.

[005] In one embodiment, a composition is provided that comprises a protein in crystalline form wherein the protein has 55%, 65%, 78%, 85%, 90%, 95%, 97%, 99% or greater identity with residues 16-314 of SEQ. ID No. 1.

[006] In one variation, the protein has activity characteristic of IspA. For example, the protein may optionally be inhibited by inhibitors of the E. coli form of IspA.

[007] The protein crystal may also diffract X-rays for a determination of structure coordinates to a resolution of 4Å, 3.5Å, 3.0Å or less.

[008] In one variation, the protein crystal has a crystal lattice in a P4<sub>1</sub>22 space group. The protein crystal may also have a crystal lattice having unit cell dimensions, +/- 5%, of a=88.80Å, b=88.80Å and c=174.99Å;  $\alpha=\beta=\gamma=90^\circ$ .

[009] In one variation, the protein has activity characteristic of IspA. For example, the protein may optionally be inhibited by inhibitors of the E. coli or human forms of IspA.

[0010] The present invention is also directed to crystallizing IspA. The present invention is also directed to the conditions useful for crystallizing IspA. It should be recognized that a wide variety of crystallization methods can be used in combination with the crystallization conditions to form crystals comprising IspA including, but not limited to, vapor diffusion, batch, dialysis, and other methods of contacting the protein solution for the purpose of crystallization.

[0011] In one embodiment, a method is provided for forming crystals of a protein comprising: forming a crystallization volume comprising: a protein that has at least 55% identity with residues 16-314 of SEQ. ID No. 1 in a concentration between 1 mg/ml and 50 mg/ml; 5-50% w/v of precipitant wherein the precipitant comprises one or more members of the group consisting of PEG MME having a molecular weight range between 300-10000, and PEG having a molecular weight range between 100-10000; optionally 0.05 to 2.5M additives wherein the additives comprise a monovalent and/or divalent salt (for example, sodium, lithium, magnesium, calcium, and the like); and storing the crystallization volume under conditions suitable for crystal formation. The method also optionally further includes performing the crystallization at a temperature between 1°C - 37°C.

[0012] The method may optionally further comprise forming a protein crystal that has a crystal lattice in a P4<sub>1</sub>22 space group. The method also optionally further comprises forming a protein crystal that has a crystal lattice having unit cell dimensions, +/- 5%, of a=88.80Å,

b=88.80Å and c=174.99Å;  $\alpha=\beta=\gamma=90$ . The invention also relates to protein crystals formed by these methods.

[0013] The present invention is also directed to structure coordinates for IspA as well as structure coordinates that are comparatively similar to these structure coordinates. It is noted that these comparatively similar structure coordinates may encompass proteins with similar sequences and/or structures, such as other IspA homologs. For example, machine-readable data storage media is provided having data storage material encoded with machine-readable data that comprises structure coordinates that are comparatively similar to the structure coordinates of IspA. The present invention is also directed to a machine readable data storage medium having data storage material encoded with machine readable data, which, when read by an appropriate machine, can display a three dimensional representation of all or a portion of a structure of IspA or a model that is comparatively similar to the structure of all or a portion of IspA.

[0014] Various embodiments of machine readable data storage medium are provided that comprise data storage material encoded with machine readable data. The machine readable data comprises: structure coordinates that have a root mean square deviation equal to or less than the RMSD value specified in Columns 3, 4 or 5 of Table 1 when compared to the structure coordinates of Figure 3, the root mean square deviation being calculated such that the portion of amino acid residues specified in Column 2 of Table 1 of each set of structure coordinates are superimposed and the root mean square deviation is based only on those amino acid residues in the structure coordinates that are also present in the portion of the protein specified in specified in Column 1 of Table 1. The amino acids being overlayed and compared need not to be identical when the RMSD calculation is performed on alpha carbons and main chain atoms but the amino acids being overlayed and compared must have identical side chains when the RMSD calculation is performed on all non-hydrogen atoms.

[0015] For example, in one embodiment where the comparison is based on the 4 Angstrom set of amino acid residues (Column 1) and is based on superimposing alpha-carbon atoms (Column 2), the structure coordinates may have a root mean square deviation equal to or less than 0.27Å, 0.18Å or 0.13Å when compared to the structure coordinates of Figure 3.

TABLE 1

AA RESIDUES TO USE TO PERFORM RMSD COMPARISON	PORTION OF EACH AA RESIDUE USED TO PERFORM RMSD COMPARISON	RMSD VALUE LESS THAN OR EQUAL TO		
IPP BINDING POCKET				
Table 2 (4 Angstrom set)	alpha-carbon atoms <sup>1</sup>	0.27	0.18	0.13
	main-chain atoms <sup>1</sup>	0.27	0.19	0.13
	all non-hydrogen <sup>2</sup>	0.57	0.37	0.28
Table 3 (7 Angstrom set)	alpha-carbon atoms <sup>1</sup>	0.90	0.60	0.45
	main-chain atoms <sup>1</sup>	0.92	0.61	0.46
	all non-hydrogen <sup>2</sup>	1.06	0.70	0.53
Table 4 (10 Angstrom set)	alpha-carbon atoms <sup>1</sup>	1.63	1.08	0.82
	main-chain atoms <sup>1</sup>	1.61	1.06	0.81
	all non-hydrogen <sup>2</sup>	1.02	0.67	0.51
RISEDRONATE BINDING POCKET				
Table 5 (4 Angstrom set)	alpha-carbon atoms <sup>1</sup>	1.63	1.08	0.82
	main-chain atoms <sup>1</sup>	1.61	1.06	0.81
	all non-hydrogen <sup>2</sup>	1.02	0.67	0.51
Table 6 (7 Angstrom set)	alpha-carbon atoms <sup>1</sup>	1.63	1.08	0.82
	main-chain atoms <sup>1</sup>	1.61	1.06	0.81
	all non-hydrogen <sup>2</sup>	1.02	0.67	0.51
Table 7 (10 Angstrom set)	alpha-carbon atoms <sup>1</sup>	1.63	1.08	0.82
	main-chain atoms <sup>1</sup>	1.61	1.06	0.81
	all non-hydrogen <sup>2</sup>	1.02	0.67	0.51
ENTIRE PROTEIN				
16-314 of SEQ. ID No. 1	alpha-carbon atoms <sup>1</sup>	1.61	1.06	0.81
	main-chain atoms <sup>1</sup>	1.59	1.05	0.80
	all non-hydrogen <sup>2</sup>	1.52	1.00	0.76

<sup>1</sup>- the RMSD computed between the atoms of all amino acids that are common to both the target and the reference in the aligned and superposed structure. The amino acids need not to be identical.

<sup>2</sup>- the RMSD computed only between identical amino acids, which are common to both the target and the reference in the aligned and superposed structure.

[0016] The present invention is also directed to a three-dimensional structure of all or a portion of IspA. This three-dimensional structure may be used to identify binding sites, to provide mutants having desirable binding properties, and ultimately, to design, characterize, or

identify ligands capable of interacting with IspA. Ligands that interact with IspA may be any type of atom, compound, protein or chemical group that binds to or otherwise associates with the protein. Examples of types of ligands include natural substrates for IspA, inhibitors of IspA, and heavy atoms. The inhibitors of IspA may optionally be used as drugs to treat therapeutic indications by modifying the in vivo activity of IspA.

[0017] In various embodiments, methods are provided for displaying a three dimensional representation of a structure of a protein comprising:

taking machine readable data comprising structure coordinates that have a root mean square deviation equal to or less than the RMSD value specified in Columns 3, 4 or 5 of Table 1 when compared to the structure coordinates of Figure 3, the root mean square deviation being calculated such that the portion of amino acid residues specified in Column 2 of Table 1 of each set of structure coordinates are superimposed and the root mean square deviation is based only on those amino acid residues in the structure coordinates that are also present in the portion of the protein specified in specified in Column 1 of Table 1;

computing a three dimensional representation of a structure based on the structure coordinates; and

displaying the three dimensional representation.

[0018] The present invention is also directed to a method for solving a three-dimensional crystal structure of a target protein using the structure of IspA.

[0019] In various embodiments, computational methods are provided comprising: taking machine readable data comprising structure coordinates that have a root mean square deviation equal to or less than the RMSD value specified in Columns 3, 4 or 5 of Table 1 when compared to the structure coordinates of Figure 3, the root mean square deviation being calculated such that the portion of amino acid residues specified in Column 2 of Table 1 of each set of structure coordinates are superimposed and the root mean square deviation is based only on those amino acid residues in the structure coordinates that are also present in the portion of the protein specified in specified in Column 1 of Table 1;

computing phases based on the structural coordinates;

computing an electron density map based on the computed phases; and

determining a three-dimensional crystal structure based on the computed electron density map.

[0020] In various embodiments, computational methods are provided comprising: taking an X-ray diffraction pattern of a crystal of the target protein; and computing a three-dimensional electron density map from the X-ray diffraction pattern by molecular replacement, wherein structure coordinates used as a molecular replacement model comprise structure coordinates that have a root mean square deviation equal to or less than the RMSD value specified in Columns 3, 4 or 5 of Table 1 when compared to the structure coordinates of Figure 3, the root mean square deviation being calculated such that the portion of amino acid residues specified in Column 2 of Table 1 of each set of structure coordinates are superimposed and the root mean square deviation is based only on those amino acid residues in the structure coordinates that are also present in the portion of the protein specified in specified in Column 1 of Table 1.

[0021] These methods may optionally further comprise determining a three-dimensional crystal structure based upon the computed three-dimensional electron density map.

[0022] The present invention is also directed to using a crystal structure of IspA, in particular the structure coordinates of IspA and the surface contour defined by them, in methods for screening, designing, or optimizing molecules or other chemical entities that interact with and preferably inhibit IspA.

[0023] One skilled in the art will appreciate the numerous uses of the inventions described herein, particularly in the areas of drug design, screening and optimization of drug candidates, as well as in determining additional unknown crystal structures. For example, a further aspect of the present invention relates to using a three-dimensional crystal structure of all or a portion of IspA and/or its structure coordinates to evaluate the ability of entities to associate with IspA. The entities may be any entity that may function as a ligand and thus may be any type of atom, compound, protein (such as antibodies) or chemical group that can bind to or otherwise associate with a protein.

[0024] In various embodiments, methods are provided for evaluating a potential of an entity to associate with a protein comprising:

creating a computer model of a protein structure using structure coordinates that comprise structure coordinates that have a root mean square deviation equal to or less than the RMSD value specified in Columns 3, 4 or 5 of Table 1 when compared to the structure coordinates of Figure 3, the root mean square deviation being calculated such that the portion

of amino acid residues specified in Column 2 of Table 1 of each set of structure coordinates are superimposed and the root mean square deviation is based only on those amino acid residues in the structure coordinates that are also present in the portion of the protein specified in specified in Column 1 of Table 1;

performing a fitting operation between the entity and the computer model; and

analyzing results of the fitting operation to quantify an association between the entity and the model.

[0025] In other embodiments, methods are provided for identifying entities that can associate with a protein comprising: generating a three-dimensional structure of a protein using structure coordinates that comprise structure coordinates that have a root mean square deviation equal to or less than the RMSD value specified in Columns 3, 4 or 5 of Table 1 when compared to the structure coordinates of Figure 3, the root mean square deviation being calculated such that the portion of amino acid residues specified in Column 2 of Table 1 of each set of structure coordinates are superimposed and the root mean square deviation is based only on those amino acid residues in the structure coordinates that are also present in the portion of the protein specified in specified in Column 1 of Table 1; and

employing the three-dimensional structure to design or select an entity that can associate with the protein; and contacting the entity with a protein having at least 55% identity with residues 16-314 of SEQ. ID No. 1.

[0026] In other embodiments, methods are provided for identifying entities that can associate with a protein comprising:

generating a three-dimensional structure of a protein using structure coordinates that comprise structure coordinates that have a root mean square deviation equal to or less than the RMSD value specified in Columns 3, 4 or 5 of Table 1 when compared to the structure coordinates of Figure 3, the root mean square deviation being calculated such that the portion of amino acid residues specified in Column 2 of Table 1 of each set of structure coordinates are superimposed and the root mean square deviation is based only on those amino acid residues in the structure coordinates that are also present in the portion of the protein specified in specified in Column 1 of Table 1; and

employing the three-dimensional structure to design or select an entity that can associate with the protein.

**[0027]** In other embodiments, methods are provided for identifying entities that can associate with a protein comprising:

computing a computer model for a protein binding pocket, at least a portion of the computer model having a surface contour that has a root mean square deviation equal to or less than a given RMSD value specified in Columns 3, 4 or 5 of Table 1 when the coordinates used to compute the surface contour are compared to the structure coordinates of Figure 3, wherein (a) the root mean square deviation is calculated by the calculation method set forth herein, (b) the portion of amino acid residues associated with the given RMSD value in Table 1 (specified in Column 2 of Table 1) are superimposed according to the RMSD calculation, and (c) the root mean square deviation is calculated based only on those amino acid residues present in both the protein being modeled and the portion of the protein associated with the given RMSD in Table 1 (specified in Column 1 of Table 1);

employing the computer model to design or select an entity that can associate with the protein; and contacting the entity with a protein having at least 55% identity with residues 16-314 of SEQ. ID No. 1.

**[0028]** In other embodiments, methods are provided for identifying entities that can associate with a protein comprising:

computing a computer model for a protein binding pocket, at least a portion of the computer model having a surface contour that has a root mean square deviation equal to or less than a given RMSD value specified in Columns 3, 4 or 5 of Table 1 when the coordinates used to compute the surface contour are compared to the structure coordinates of Figure 3, wherein (a) the root mean square deviation is calculated by the calculation method set forth herein, (b) the portion of amino acid residues associated with the given RMSD value in Table 1 (specified in Column 2 of Table 1) are superimposed according to the RMSD calculation, and (c) the root mean square deviation is calculated based only on those amino acid residues present in both the protein being modeled and the portion of the protein associated with the given RMSD in Table 1 (specified in Column 1 of Table 1); and

employing the computer model to design or select an entity that can associate with the protein.

**[0029]** In other embodiments, methods are provided for evaluating the ability of an entity to associate with a protein, the method comprising:



constructing a computer model defined by structure coordinates that have a root mean square deviation equal to or less than the RMSD value specified in Columns 3, 4 or 5 of Table 1 when compared to the structure coordinates of Figure 3, the root mean square deviation being calculated such that the portion of amino acid residues specified in Column 2 of Table 1 of each set of structure coordinates are superimposed and the root mean square deviation is based only on those amino acid residues in the structure coordinates that are also present in the portion of the protein specified in specified in Column 1 of Table 1; and

selecting an entity to be evaluated by a method selected from the group consisting of (i) assembling molecular fragments into the entity, (ii) selecting an entity from a small molecule database, (iii) *de novo* ligand design of the entity, and (iv) modifying a known ligand for IspA, or a portion thereof; performing a fitting program operation between computer models of the entity to be evaluated and the binding pocket in order to provide an energy-minimized configuration of the entity in the binding pocket; and evaluating the results of the fitting operation to quantify the association between the entity and the binding pocket model in order to evaluate the ability of the entity to associate with the binding pocket.

[0030] In other embodiments, methods are provided for evaluating the ability of an entity to associate with a protein, the method comprising:

computing a computer model for a protein binding pocket, at least a portion of the computer model having a surface contour that has a root mean square deviation equal to or less than a given RMSD value specified in Columns 3, 4 or 5 of Table 1 when the coordinates used to compute the surface contour are compared to the structure coordinates of Figure 3, wherein (a) the root mean square deviation is calculated by the calculation method set forth herein, (b) the portion of amino acid residues associated with the given RMSD value in Table 1 (specified in Column 2 of Table 1) are superimposed according to the RMSD calculation, and (c) the root mean square deviation is calculated based only on those amino acid residues present in both the protein being modeled and the portion of the protein associated with the given RMSD in Table 1 (specified in Column 1 of Table 1); and

selecting an entity to be evaluated by a method selected from the group consisting of (i) assembling molecular fragments into the entity, (ii) selecting an entity from a small molecule database, (iii) *de novo* ligand design of the entity, and (iv) modifying a known ligand for IspA, or a portion thereof; performing a fitting program operation between computer models of the

entity to be evaluated and the binding pocket in order to provide an energy-minimized configuration of the entity in the binding pocket; and evaluating the results of the fitting operation to quantify the association between the entity and the binding pocket model in order to evaluate the ability of the entity to associate with the binding pocket.

[0031] In regard to each of these embodiments, the protein may optionally have activity characteristic of IspA. For example, the protein may optionally be inhibited by inhibitors of the *E. coli* or human forms of IspA.

[0032] In another embodiment, a method is provided for identifying an entity that associates with a protein comprising: taking structure coordinates from diffraction data obtained from a crystal of a protein that has at least 55%, 65%, 78%, 85%, 90%, 95%, 97%, 99% or more identity with the residues 16-314 of SEQ. ID No. 1; and performing rational drug design using a three dimensional structure that is based on the obtained structure coordinates. The protein crystals may optionally have a crystal lattice having unit cell dimensions, +/- 5%, of  $a=88.80\text{\AA}$ ,  $b=88.80\text{\AA}$  and  $c=174.99\text{\AA}$ ;  $\alpha=\beta=\gamma=90$ . The method may optionally further comprise selecting one or more entities based on the rational drug design and contacting the selected entities with the protein. The method may also optionally further comprise measuring an activity of the protein when contacted with the one or more entities. The method also may optionally further comprise comparing activity of the protein in a presence of and in the absence of the one or more entities; and selecting entities where activity of the protein changes depending whether a particular entity is present. The method also may optionally further comprise contacting cells expressing the protein with the one or more entities and detecting a change in a phenotype of the cells when a particular entity is present.

#### BRIEF DESCRIPTION OF THE FIGURES

[0033] Figure 1 illustrates SEQ. ID Nos. 1 and 2 referred to in this application.

[0034] Figure 2 illustrates crystal of IspA corresponding to SEQ. ID No. 1, having a crystal lattice in a  $P4_122$  space group and unit cell dimensions, +/- 5%, of  $a=88.80\text{\AA}$ ,  $b=88.80\text{\AA}$  and  $c=174.99\text{\AA}$ ;  $\alpha=\beta=\gamma=90$ .

[0035] Figure 3 lists a set of atomic structure coordinates for IspA as derived by X-ray crystallography from a crystal that comprises the protein. The following abbreviations are used in Figure 3: "X, Y, Z" crystallographically define the atomic position of the element measured;

“B” is a thermal factor that measures movement of the atom around its atomic center; “Occ” is an occupancy factor that refers to the fraction of the molecules in which each atom occupies the position specified by the coordinates (a value of “1” indicates that each atom has the same conformation, i.e., the same position, in all molecules of the crystal).

[0036] Figure 4 illustrates a ribbon diagram overview of the structure of IspA, highlighting secondary structural elements of the protein.

[0037] Figure 5A illustrates residues within 4.0 Å of the IPP binding pocket for IspA.

[0038] Figure 5B illustrates residues within 4.0 Å of the Risedronate binding pocket for IspA.

[0039] Figure 6 illustrates a system that may be used to carry out instructions for displaying a crystal structure of IspA encoded on a storage medium.

## DETAILED DESCRIPTION OF THE INVENTION

[0040] The present invention relates to a mevalonate pathway enzyme responsible for the synthesis of farnesyl pyrophosphate (FPP), and more specifically to IspA also known as farnesyl pyrophosphate synthase (FPPS) and farnesyl diphosphate synthase (FDPS), referred to herein as IspA. Provided is IspA in crystalline form, methods of forming crystals comprising IspA, methods of using crystals comprising IspA, a crystal structure of IspA, and methods of using the crystal structure.

[0041] In describing protein structure and function herein, reference is made to amino acids comprising the protein. The amino acids may also be referred to by their conventional abbreviations; A = Ala = Alanine; T = Thr = Threonine; V = Val = Valine; C = Cys = Cysteine; L = Leu = Leucine; Y = Tyr = Tyrosine; I = Ile = Isoleucine; N = Asn = Asparagine; P = Pro = Proline; Q = Gln = Glutamine; F = Phe = Phenylalanine; D = Asp = Aspartic Acid; W = Trp = Tryptophan; E = Glu = Glutamic Acid; M = Met = Methionine; K = Lys = Lysine; G = Gly = Glycine; R = Arg = Arginine; S = Ser = Serine; and H = His = Histidine.

### 1. IspA

[0042] IspA, also known as farnesyl pyrophosphate synthase (FPPS) and farnesyl diphosphate synthase (FDPS), is a mevalonate pathway enzyme responsible for the synthesis of farnesyl pyrophosphate (FPP), a key branchpoint intermediate for the biosynthesis of

cholesterol, prenylated proteins, ubiquinones, dolichols, and heme a. The enzyme belongs to the larger family of isoprenoid diphosphate synthases that catalyze prenyl transfer to isopentyl pyrophosphate (IPP). Although different isoprenoid diphosphate synthase family members catalyze different reactions, all members of the family possess two conserved ASP rich motifs (DDXXD and DDXXXXD) crucial for catalysis. In cells, IspA synthesizes the C15 isoprenoid FPP through two sequential condensations of allylic and homoallylic isoprenoid units. In the first reaction isopentyl pyrophosphate (IPP) and dimethylallyl pyrophosphate (DMAPP) are condensed to form the C10 isoprenoid geranyl pyrophosphate (GPP). In the second reaction GPP is condensed with another molecule of IPP to form FPP, completing the C5 to C15 isoprenoid elongation.

[0043] Close homologues of the IspA enzyme are found in organism from all the three kingdoms of life, including a human enzyme that shares ~26% sequence identity. Notably, residues identified as contacting the allylic and homoallylic substrates are highly conserved among homologues indicating that the substrate-bound form of the E.coli enzyme will be useful for ligand design directed at the human enzyme.

[0044] Human FPPS is the molecular target of the N-containing bisphosphonate drugs used to treat osteoporosis. Bisphosphonates are analogs of pyrophosphate (P-O-P) in which the central pyrophosphate oxygen is replaced by a carbon with various side chains. The highly charged phosphate groups target and concentrate bisphosphonates on bone surfaces where they are absorbed by osteoclastic cells responsible for bone resorption. The biological consequences of bisphosphonate-mediated FPPS inhibition is thought to be a direct result of reduced intracellular levels of FPP and the C20 isoprenoid geranylgeranyl pyrophosphate (GGPP). These two molecules are substrates for prenyl:protein transferases that attach an isoprenoid lipid onto the C-terminus of small GTPases (such as Ras, Rac, Rho, and CDC42) to direct their subcellular localization and influence key protein:protein interactions. Disruption of these activities through bisphosphonate-mediated isoprenoid depletion disrupts the function of the osteoclast, which undergoes apoptosis, resulting in reduced bone resorption and lower bone turnover.

[0045] The bisphosphonate-mediated reduction of FPP has parallels with the statin family of drugs that inhibit HMG-CoA reductase, a mevalonate pathway enzyme that acts upstream of IspA. Statins, unlike bisphosphonates are highly lipophilic, and as such are

targeted to the liver rather than bone. In the liver, statin-mediated HMG-CoA reductase inhibition indirectly reduces IspA levels and thus decreases the rate of cholesterol biosynthesis. The beneficial effects of statins, however, are not only mediated through a reduction in serum cholesterol, but also by disruption of trimeric G-protein signaling. Thus, like bisphosphonate-mediated inhibition of FPPS, these cholesterol-independent effects of statins' result from a reduction in the synthesis of key isoprenoid intermediates. As such, bisphosphonate-mediated inhibition of FPPS and statin mediated inhibition of HMG-CoA is currently being investigated for the treatment of a diverse range of clinical indications including cell proliferation and metastases, angiogenesis, inflammation, obesity and the treatment of parasitic diseases.

[0046] In one embodiment, IspA comprises the *E. coli* form of full length IspA, set forth herein as SEQ. ID No. 1 (GenBank Accession Number NM\_ D00694).

[0047] In another embodiment, IspA comprises residues 16-314 of SEQ. ID No. 1 which comprises the active site domain of wild-type IspA that is represented in the set of structural coordinates shown in Figure 3.

[0048] It should be recognized that the invention may be readily extended to various variants of wild-type IspA and variants of fragments thereof. In another embodiment, IspA comprises a sequence that has at least 55% identity, preferably at least 65%, 78%, 85%, 90%, 95%, 97%, 99% or higher identity with SEQ. ID No. 1.

[0049] It is also noted that the above sequences of IspA are also intended to encompass isoforms, mutants and fusion proteins of these sequences.

[0050] With the crystal structure provided herein, it is now known where amino acid residues are positioned in the structure. As a result, the impact of different substitutions can be more easily predicted and understood.

[0051] For example, based on the crystal structure, applicants have determined that the IspA amino acids shown in Table 2 encompass a 4-Angstrom radius around the IspA active site and thus likely to interact with any active site inhibitor of IspA. Applicants have also determined that the amino acids of Table 3 encompass a 7-Angstrom radius around the IspA active site. Further it has been determined that the amino acids of Table 4 encompass a 10-Angstrom radius around the IspA active site. It is noted that there is one IspA molecule in the asymmetric unit, referred to as chain A. Structural coordinates appear in Figure 3. It is noted

that the sequence and structure of the residues in the active site may also be conserved and hence pertinent to other homologues of IspA.

**[0052]** One or more of the sets of amino acids set forth in the tables is preferably conserved in a variant of IspA. Hence, IspA may optionally comprise a sequence that has at least 55% identity, preferably at least 65%, 78%, 85%, 90%, 95%, 97%, 99% or higher identity with any one of the above sequences (e.g., all of SEQ. ID No. 1 or residues 16-314 of SEQ. ID No. 1) where at least the residues shown in Tables 2, 3, 4, 5, 6 and/or 7 are conserved with the exception of 0, 1, 2, 3, or 4 residues. It should be recognized that one might optionally vary some of the binding site residues in order to determine the effect such changes have on structure or activity.

**Table 2:** Amino Acids encompassed by a 4-Angstrom radius around the IPP binding pocket.

GLY	65	LYS	66	ARG	69	HIS	98
LEU	102	ARG	117	THR	203	PHE	240
GLN	241	ASP	244	ARG	318		

**Table 3:** Amino Acids encompassed by a 7-Angstrom radius around the IPP binding pocket.

LEU	63	GLY	64	GLY	65	LYS	66	ARG	67
LEU	68	ARG	69	GLU	95	HIS	98	ALA	99
SER	101	LEU	102	ASP	105	ARG	116	ARG	117
LYS	202	THR	203	GLY	204	LEU	206	ILE	207
PHE	240	GLN	241	ASP	244	ASP	245	LEU	247
LEU	256	LYS	258	ARG	318	LYS	320		

**Table 4:** Amino Acids encompassed by a 10 -Angstrom radius around the IPP binding pocket.

LEU	28	VAL	32	ASN	36	LEU	39	TYR	59
GLY	60	ALA	61	LEU	62	LEU	63	GLY	64
GLY	65	LYS	66	ARG	67	LEU	68	ARG	69
PRO	70	PHE	71	LEU	72	VAL	94	GLU	95
CYS	96	ILE	97	HIS	98	ALA	99	TYR	100
SER	101	LEU	102	ILE	103	ASP	105	ASP	106
ASP	111	LEU	115	ARG	116	ARG	117	GLY	118
LEU	119	MET	175	GLN	179	ASP	182	ILE	198
HIS	199	ARG	200	HIS	201	LYS	202	THR	203

GLY	204	ALA	205	LEU	206	ILE	207	ILE	236
GLY	237	LEU	238	ALA	239	PHE	240	GLN	241
VAL	242	GLN	243	ASP	244	ASP	245	ILE	246
LEU	247	ASP	248	THR	255	LEU	256	GLY	257
LYS	258	ASP	263	LEU	311	TYR	314	ILE	315
ARG	318	ASN	319	LYS	320				

**Table 5:** Amino Acids encompassed by a 4-Angstrom radius around the Risedronate binding pocket.

SER	101	LEU	102	ASP	105	ASP	111	ARG	116
GLN	179	LYS	202	THR	203	GLN	241	ASP	244
LYS	258								

**Table 6:** Amino Acids encompassed by a 7-Angstrom radius around the Risedronate binding pocket.

ARG	69	HIS	98	SER	101	LEU	102	ASP	105
ASP	106	ASP	111	ASP	113	LEU	115	ARG	116
MET	175	GLY	178	GLN	179	ASP	182	ILE	198
HIS	199	HIS	201	LYS	202	THR	203	GLY	204
PHE	240	GLN	241	ASP	244	ASP	245	ASP	248
LYS	258	ARG	259	ALA	262	ASP	263	LYS	268

**Table 7:** Amino Acids encompassed by a 10-Angstrom radius around the Risedronate binding pocket.

LYS	66	ARG	69	ILE	97	HIS	98	ALA	99
TYR	100	SER	101	LEU	102	ILE	103	HIS	104
ASP	105	ASP	106	LEU	107	MET	110	ASP	111
ASP	112	ASP	113	ASP	114	LEU	115	ARG	116
ARG	117	THR	121	ALA	169	SER	170	GLY	171
GLY	174	MET	175	CYS	176	GLY	177	GLY	178
GLN	179	ALA	180	LEU	181	ASP	182	LEU	183
GLU	186	LEU	195	ILE	198	HIS	199	ARG	200
HIS	201	LYS	202	THR	203	GLY	204	ALA	205
LEU	206	ILE	207	GLY	237	PHE	240	GLN	241
VAL	242	GLN	243	ASP	244	ASP	245	ILE	246
LEU	247	ASP	248	VAL	249	LEU	256	GLY	257
LYS	258	ARG	259	GLN	260	GLY	261	ALA	262
ASP	263	GLN	264	LEU	266	LYS	268	SER	269
THR	270	TYR	271	PRO	272	ARG	318	LYS	320

[0053] With the benefit of the crystal structure and guidance provided by Tables 2, 3, 4, 5, 6 and 7, a wide variety of IspA variants (e.g., insertions, deletions, substitutions, etc.) that fall within the above specified identity ranges may be designed and manufactured utilizing recombinant DNA techniques well known to those skilled in the art, particularly in view of the knowledge of the crystal structure provided herein. These modifications can be used in a number of combinations to produce the variants. The present invention is useful for crystallizing and then solving the structure of the range of variants of IspA.

[0054] Variants of IspA may be insertional variants in which one or more amino acid residues are introduced into a predetermined site in the IspA sequence. For instance, insertional variants can be fusions of heterologous proteins or polypeptides to the amino or carboxyl terminus of the subunits.

[0055] Variants of IspA also may be substitutional variants in which at least one residue has been removed and a different residue inserted in its place. Non-natural amino acids (i.e. amino acids not normally found in native proteins), as well as isosteric analogs (amino acid or otherwise) may optionally be employed in substitutional variants. Examples of suitable substitutions are well known in the art, such as the Glu→Asp, Asp→Glu, Ser→Cys, and Cys→Ser for example.

[0056] Another class of variants is deletional variants, which are characterized by the removal of one or more amino acid residues from the IspA sequence.

[0057] Other variants may be produced by chemically modifying amino acids of the native protein (e.g., diethylpyrocarbonate treatment that modifies histidine residues). Preferred are chemical modifications that are specific for certain amino acid side chains. Specificity may also be achieved by blocking other side chains with antibodies directed to the side chains to be protected. Chemical modification includes such reactions as oxidation, reduction, amidation, deamidation, or substitution of bulky groups such as polysaccharides or polyethylene glycol.

[0058] Exemplary modifications include the modification of lysinyl and amino terminal residues by reaction with succinic or other carboxylic acid anhydrides. Modification with these agents has the effect of reversing the charge of the lysinyl residues. Other suitable reagents for modifying amino-containing residues include imidoesters such as methyl picolinimide; pyridoxal phosphate; pyridoxal chloroborohydride; trinitrobenzenesulfonic acid; O-



methylisourea, 2,4-pentanedione; and transaminaseN catalyzed reaction with glyoxylate, and N-hydroxysuccinamide esters of polyethylene glycol or other bulky substitutions.

[0059] Arginyl residues may be modified by reaction with a number of reagents, including phenylglyoxal, 2,3-butanedione, 1,2-cyclohexanedione, and ninhydrin. Modification of arginine residues requires that the reaction be performed in alkaline conditions because of the high  $pK_a$  of the guanidine functional group. Furthermore, these reagents may react with the groups of lysine as well as the arginine epsilon-amino group.

[0060] Tyrosyl residues may also be modified to introduce spectral labels into tyrosyl residues by reaction with aromatic diazonium compounds or tetranitromethane, forming O-acetyl tyrosyl species and 3-nitro derivatives, respectively. Tyrosyl residues may also be iodinated using  $^{125}\text{I}$  or  $^{131}\text{I}$  to prepare labeled proteins for use in radioimmunoassays.

[0061] Carboxyl side groups (aspartyl or glutamyl) may be selectively modified by reaction with carbodiimides or they may be converted to asparaginyl and glutaminyl residues by reaction with ammonium ions. Conversely, asparaginyl and glutaminyl residues may be deamidated to the corresponding aspartyl or glutamyl residues, respectively, under mildly acidic conditions. Either form of these residues falls within the scope of this invention.

[0062] Other modifications that may be formed include the hydroxylation of proline and lysine, phosphorylation of hydroxyl groups of seryl or threonyl groups of lysine, arginine and histidine side chains (T. E. Creighton, *Proteins: Structure and Molecular Properties*, W.H. Freeman & Co., San Francisco, pp. 79-86, 1983), acetylation of the N-terminal amine and amidation of any C-terminal carboxyl group.

[0063] As can be seen, modifications of the nucleic sequence encoding IspA may be accomplished by a variety of well-known techniques, such as site-directed mutagenesis (see, Gillman and Smith, *Gene* 8:81-97 (1979) and Roberts, S. *et al.*, *Nature* 328:731-734 (1987)). When modifications are made, these modifications may optionally be evaluated for their effect on a variety of different properties including, for example, solubility, crystallizability and a modification to the protein's structure and activity.

[0064] In one variation, the variant and/or fragment of E. coli IspA is functional in the sense that the resulting protein is capable of associating with at least one same chemical entity that is also capable of selectively associating with a protein comprising the E. coli IspA (e.g.,

residues 16-314 of SEQ. ID No. 1) since this common associative ability evidences that at least a portion of the native structure has been conserved.

[0065] It is noted the activity of the native protein need not necessarily be conserved. Rather, amino acid substitutions, additions or deletions that interfere with native activity but which do not significantly alter the three-dimensional structure of the domain are specifically contemplated by the invention. Crystals comprising such variants of IspA, and the atomic structure coordinates obtained there from, can be used to identify compounds that bind to the native domain. These compounds may affect the activity of the native domain.

[0066] Amino acid substitutions, deletions and additions that do not significantly interfere with the three-dimensional structure of IspA will depend, in part, on the region where the substitution, addition or deletion occurs in the crystal structure. These modifications to the protein can now be made far more intelligently with the crystal structure information provided herein. In highly variable regions of the molecule, non-conservative substitutions as well as conservative substitutions may be tolerated without significantly disrupting the three-dimensional structure of the molecule. In highly conserved regions, or regions containing significant secondary structure, conservative amino acid substitutions are preferred.

[0067] Conservative amino acid substitutions are well known in the art, and include substitutions made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity and/or the amphipathic nature of the amino acid residues involved. For example, negatively charged amino acids include aspartic acid and glutamic acid; positively charged amino acids include lysine and arginine; amino acids with uncharged polar head groups having similar hydrophilicity values include the following: leucine, isoleucine, valine; glycine, alanine; asparagine, glutamine; serine, threonine; phenylalanine, tyrosine. Other conservative amino acid substitutions are well known in the art.

[0068] It should be understood that the protein may be produced in whole or in part by chemical synthesis. As a result, the selection of amino acids available for substitution or addition is not limited to the genetically encoded amino acids. Indeed, mutants may optionally contain non-genetically encoded amino acids. Conservative amino acid substitutions for many of the commonly known non-genetically encoded amino acids are well known in the art. Conservative substitutions for other amino acids can be determined based on their physical properties as compared to the properties of the genetically encoded amino acids.

[0069] In some instances, it may be particularly advantageous or convenient to substitute, delete and/or add amino acid residues in order to provide convenient cloning sites in cDNA encoding the polypeptide, to aid in purification of the polypeptide, etc. Such substitutions, deletions and/or additions which do not substantially alter the three dimensional structure of IspA will be apparent to those having skills in the art, particularly in view of the three dimensional structure of IspA provided herein.

## 2. Cloning, Expression and Purification of IspA

[0070] The gene encoding IspA can be isolated from RNA, cDNA or cDNA libraries. In this case, the portion of the gene encoding amino acid residues 16-314 (SEQ. ID No. 1), corresponding to E. coli IspA, was isolated and is shown as SEQ. ID No. 2.

[0071] Construction of expression vectors and recombinant proteins from the DNA sequence encoding IspA may be performed by various methods well known in the art. For example, these techniques may be performed according to Sambrook et al., Molecular Cloning-A Laboratory Manual, Cold Spring Harbor, N.Y. (1989), and Kriegler, M., Gene Transfer and Expression, A Laboratory Manual, Stockton Press, New York (1990).

[0072] A variety of expression systems and hosts may be used for the expression of IspA. Example 1 provides one such expression system.

[0073] Once expressed, purification steps are employed to produce IspA in a relatively homogeneous state. In general, a higher purity solution of a protein increases the likelihood that the protein will crystallize. Typical purification methods include the use of centrifugation, partial fractionation, using salt or organic compounds, dialysis, conventional column chromatography, (such as ion exchange, molecular sizing chromatography, etc.), high performance liquid chromatography (HPLC), and gel electrophoresis methods (see, e.g., Deutcher, "Guide to Protein Purification" in Methods in Enzymology (1990), Academic Press, Berkeley, California).

## 3. Crystallization and Crystals Comprising IspA

[0074] One aspect of the present invention relates to methods for forming crystals comprising IspA as well as crystals comprising IspA.

[0075] In one embodiment, a method for forming crystals comprising IspA is provided comprising forming a crystallization volume comprising IspA, one or more precipitants, optionally a buffer, optionally a monovalent and/or divalent salt and optionally an organic solvent; and storing the crystallization volume under conditions suitable for crystal formation.

[0076] In yet another embodiment, a method for forming crystals comprising IspA is provided comprising forming a crystallization volume comprising IspA in solution comprising the components shown in Table 8; and storing the crystallization volume under conditions suitable for crystal formation.

**Table 8**

**Precipitant**

5-65% w/v of precipitant wherein the precipitant comprises one or more members of the group consisting of PEG MME having a molecular weight range between 500-10000, PEG having a molecular weight range between 100-10000, MPD, and ethanol. 0.3-2.0M Sodium, potassium or ammonium phosphate.

**pH**

pH 4-10. Buffers that may be used include, but are not limited to tris, bicine, phosphate, cacodylate, acetate, citrate, HEPES, PIPES, MES and combinations thereof.

**Additives**

Optionally 0.05 to 2.5M additives wherein the additives comprise a monovalent and/or divalent salt (for example, sodium, lithium, magnesium, calcium, and the like)

***Protein Concentration***

1 mg/ml - 50 mg/ml

**Temperature**

1°C - 25°C

[0077] In yet another embodiment, a method for forming crystals comprising IspA is provided comprising forming a crystallization volume comprising IspA; introducing crystals comprising IspA as nucleation sites, and storing the crystallization volume under conditions suitable for crystal formation.

[0078] Crystallization experiments may optionally be performed in volumes commonly used in the art, for example typically 15, 10, 5, 2 microliters or less. It is noted that the crystallization volume optionally has a volume of less than 1 microliter, optionally 500, 250, 150, 100, 50 or less nanoliters.

[0079] It is also noted that crystallization may be performed by any crystallization method including, but not limited to batch, dialysis and vapor diffusion (e.g., sitting drop and hanging drop) methods. Micro, macro and/or streak seeding of crystals may also be performed to facilitate crystallization.

[0080] It should be understood that forming crystals comprising IspA and crystals comprising IspA according to the invention are not intended to be limited to the E. coli form of IspA shown in SEQ. ID No. 1, fragments comprising residues 16-314 of SEQ. ID No. 1 and fragments comprising residues 16-314 of SEQ. ID No. 1. Rather, it should be recognized that the invention may be extended to various other fragments and variants of wild-type IspA as described above.

[0081] It should also be understood that forming crystals comprising IspA and crystals comprising IspA according to the invention may be such that IspA is optionally complexed with one or more ligands and one or more copies of the same ligand. The ligand used to form the complex may be any ligand capable of binding to IspA. In one variation, the ligand is a natural substrate. In another variation, the ligand is an inhibitor.

[0082] In one particular embodiment, IspA crystals have a crystal lattice in the P4<sub>1</sub>22 space group. IspA crystals may also optionally have unit cell dimensions, +/- 5%, of a=88.80Å, b=88.80Å and c=174.99Å;  $\alpha=\beta=\gamma=90^\circ$ . IspA crystals also preferably are capable of diffracting X-rays for determination of atomic coordinates to a resolution of 4Å, 3.5Å, 3.0Å or better.

[0083] Crystals comprising IspA may be formed by a variety of different methods known in the art. For example, crystallizations may be performed by batch, dialysis, and vapor diffusion (sitting drop and hanging drop) methods. A detailed description of basic protein crystallization setups may be found in McRee, D., Practical Protein Crystallography, 2<sup>nd</sup> Ed. (1999), Academic Press Inc. Further descriptions regarding performing crystallization experiments are provided in Stevens, et al. (2000) *Curr. Opin. Struct. Biol.*: 10(5):558-63, and U.S. Patent Nos. 6,296,673, 5,419,278, and 5,096, 676.

[0084] In one variation, crystals comprising IspA are formed by mixing substantially pure IspA with an aqueous buffer containing a precipitant at a concentration just below a concentration necessary to precipitate the protein. One suitable precipitant for crystallizing IspA is polyethylene glycol (PEG), which combines some of the characteristics of the salts and other organic precipitants (see, for example, Ward et al., J. Mol. Biol. 98:161, 1975, and McPherson, J. Biol. Chem. 251:6300, 1976).

[0085] During a crystallization experiment, water is removed by diffusion or evaporation to increase the concentration of the precipitant, thus creating precipitating conditions for the protein. In one particular variation, crystals are grown by vapor diffusion in hanging drops or sitting drops. According to these methods, a protein/precipitant solution is formed and then allowed to equilibrate in a closed container with a larger aqueous reservoir having a precipitant concentration for producing crystals. The protein/precipitant solution continues to equilibrate until crystals grow.

[0086] By performing submicroliter volume sized crystallization experiments, as detailed in U.S. Patent No. 6,296,673, effective crystallization conditions for forming crystals of a IspA complex were obtained. In order to accomplish this, systematic broad screen crystallization trials were performed on an IspA complex using the sitting drop technique. Over 1000 individual trials were performed in which pH, temperature and precipitants were varied. In each experiment, a 100nL mixture of IspA complex and precipitant was placed on a platform positioned over a well containing 100μL of the precipitating solution. Precipitate and crystal formation was detected in the sitting drops. Fine screening was then carried out for those crystallization conditions that appeared to produce precipitate and/or crystal in the drops.

[0087] Based on the crystallization experiments that were performed, a thorough understanding of how different crystallization conditions affect IspA crystallization was obtained. Based on this understanding, a series of crystallization conditions were identified that may be used to form crystals comprising IspA. These conditions are summarized in Table 8. A particular example of crystallization conditions that may be used to form diffraction quality crystals of the IspA complex is detailed in Example 2. Figure 2 illustrates crystals of the IspA complex formed using the crystallization conditions provided in Table 8.

[0088] One skilled in the art will recognize that the crystallization conditions provided in Table 8 and Example 2 can be varied and still yield protein crystals comprising IspA. For

example, it is noted that variations on the crystallization conditions described herein can be readily determined by taking the conditions provided in Table 8 and performing fine screens around those conditions by varying the type and concentration of the components in order to determine additional suitable conditions for crystallizing IspA, variants of IspA, and ligand complexes thereof.

[0089] Crystals comprising IspA have a wide range of uses. For example, now that crystals comprising IspA have been produced, it is noted that crystallizations may be performed using such crystals as a nucleation site within a concentrated protein solution. According to this variation, a concentrated protein solution is prepared and crystalline material (microcrystals) is used to 'seed' the protein solution to assist nucleation for crystal growth. If the concentrations of the protein and any precipitants are optimal for crystal growth, the seed crystal will provide a nucleation site around which a larger crystal forms. Given the ability to form crystals comprising IspA according to the present invention, the crystals so formed can be used by this crystallization technique to initiate crystal growth of other IspA comprising crystals, including IspA complexed to other ligands.

[0090] As will be described herein in greater detail, crystals may also be used to perform X-ray or neutron diffraction analysis in order to determine the three-dimensional structure of IspA and, in particular, to assist in the identification of its active site. Knowledge of the binding site region allows rational design and construction of ligands including inhibitors. Crystallization and structural determination of IspA mutants having altered bioactivity allows the evaluation of whether such changes are caused by general structure deformation or by side chain alterations at the substitution site.

#### **4. X-Ray Data Collection and Structure Determination**

[0091] Crystals comprising IspA may be obtained as described above in Section 3. As described herein, these crystals may then be used to perform X-ray data collection and for structure determination.

[0092] In one embodiment, described in Example 2, crystals of IspA were obtained where IspA has the sequence of residues shown in SEQ. ID No. 1. These particular crystals were used to determine the three dimensional structure of IspA. However, it is noted that other crystals comprising IspA including different IspA variants, fragments, and complexes thereof may also be used.

[0093] Diffraction data was collected from cryocooled crystals (100K) of IspA\_Ec at the Advanced Light Source beam line 5.0.3 using an ADSC CCD detector. The diffraction pattern of IspA displayed symmetry consistent with a  $P4_122$  space group, with unit cell dimensions  $a=88.80 \text{ \AA}$ ,  $b=88.00 \text{ \AA}$  and  $c=174.99 \text{ \AA}$ . Data were collected and integrated to  $1.95 \text{ \AA}$  with DENZO and scaled with SCALEPACK.

[0094] All crystallographic calculations were performed using the CCP4 program package (Collaborative Computational Project, N. The CCP4 Suite: Programs for Protein Crystallography. *Acta Cryst.* D50, 760-763 (1994)). The initial phases for IspA\_Ec were obtained by the molecular replacement method using the program AMORE. The coordinates of an unliganded IspA from *S. Aureus* (NCBI accession code – NP\_646291), which were determined at Syrrx using MAD techniques, were used as a search model. The highest solution from the translational function was subjected to a rigid body rotation followed by positional refinement against the maximum likelihood method as implemented in REFMAC5 (CCP4). Overall refinement employed iterative rounds of manual model building with Xfit (McRee, D.E. XtalView/Xfit-A versatile program for manipulating atomic coordinates and electron density *J. Struct. Biol.* 125, 156-65 (1999)) followed by positional refinement with REFMAC5 (CCP4). All stages of model refinement were carried with bulk solvent correction and anisotropic scaling. The data collection and data refinement statistics are given in Table 9.

TABLE 9

<b>Data collection</b>		IspA_Ec
X-ray source		ALS-BL5.0.3
Wavelength [ $\text{\AA}$ ]		1.0
Resolution [ $\text{\AA}$ ]		87-1.95
Observations (unique)		51703
Redundancy		7.6
Completeness	overall (outer shell)	99.3% (96.2%)
$I/\sigma(I)$	overall (outer shell)	27.8 (3.0)
$R_{\text{symm}}$	overall (outer shell)	7.7% (51.6%)
<b>Refinement</b>		
Reflections used		49037
R-factor		20.6%
$R_{\text{free}}$		23.9%
r.m.s bonds		0.010



r.m.s angles	1.23
$R_{\text{symm}} = \frac{\sum_{hkl} \sum_i  I(hkl)_i - \langle I(hkl) \rangle }{\sum_{hkl} \sum_i \langle I(hkl) \rangle}$ over I observations of a reflection hkl	

[0095] During structure determination, it was realized that the asymmetric unit comprised two IspA molecules. Structure coordinates were determined for this complex and the resultant set of structural coordinates from the refinement are presented in Figure 3.

[0096] It is noted that the sequence of the structure coordinates presented in Figure 3 differ in some regards from the sequence shown in SEQ. ID No. 1. Structure coordinates are reported for residues 16-314.

[0097] Those of skill in the art understand that a set of structure coordinates (such as those in Figure 3) for a protein or a protein-complex or a portion thereof, is a relative set of points that define a shape in three dimensions. Thus, it is possible that an entirely different set of structure coordinates could define a similar or identical shape. Moreover, slight variations in the individual coordinates may have little effect on overall shape. In terms of binding pockets, these variations would not be expected to significantly alter the nature of ligands that could associate with those pockets. The term "binding pocket" as used herein refers to a region of the protein that, as a result of its shape, favorably associates with a ligand.

[0098] These variations in coordinates may be generated because of mathematical manipulations of the IspA structure coordinates. For example, the sets of structure coordinates shown in Figure 3 could be manipulated by crystallographic permutations of the structure coordinates, fractionalization of the structure coordinates, application of a rotation matrix, integer additions or subtractions to sets of the structure coordinates, inversion of the structure coordinates or any combination of the above.

[0099] Alternatively, modifications in the crystal structure due to mutations, additions, substitutions, and/or deletions of amino acids or other changes in any of the components that make up the crystal could also account for variations in structure coordinates. If such variations are within an acceptable standard error as compared to the original coordinates, the resulting three-dimensional shape should be considered to be the same. Thus, for example, a ligand that bound to the active site binding pocket of IspA would also be expected to bind to another binding pocket whose structure coordinates defined a shape that fell within the acceptable error.

[00100] Various computational methods may be used to determine whether a particular protein or a portion thereof (referred to here as the “target protein”), typically the binding pocket, has a high degree of three-dimensional spatial similarity to another protein (referred to here as the “reference protein”) against which the target protein is being compared.

[00101] The process of comparing a target protein structure to a reference protein structure may generally be divided into three steps: 1) defining the equivalent residues and/or atoms for the target and reference proteins, 2) performing a fitting operation between the proteins; and 3) analyzing the results. These steps are described in more detail below. All structure comparisons reported herein and the structure comparisons claimed are intended to be based on the particular comparison procedure described below.

[00102] Equivalent residues or atoms can be determined based upon an alignment of primary sequences of the proteins, an alignment of their structural domains or as a combination of both. Sequence alignments generally implement the dynamic programming algorithm of Needleman and Wunsch [*J. Mol. Biol.* 48: 442-453, 1970]. For the purpose of this invention the sequence alignment was performed using the publicly available software program MOE (Chemical Computing Group Inc.) package version 2002.3, as described in the accompanying User's Manual. When using the MOE program, alignment was performed in the sequence editor window using the ALIGN option utilizing the following program parameters: Initial pairwise Build-up: ON, Substitution Matrix: Blosum62, Round Robin: ON, Gap Start: 7, Gap Extend: 1, Iterative Refinement: ON, Build-up: TREE-BASED, Secondary Structure: NONE, Structural Alignment: ENABLED, Gap Start: 1, Gap Extend: 0.1

[00103] Once aligned, a rigid body fitting operation is performed where the structure for the target protein is translated and rotated to obtain an optimum fit relative to the structure of the reference protein. The fitting operation uses an algorithm that computes the optimum translation and rotation to be applied to the moving structure, such that the root mean square deviation of the fit over the specified pairs of equivalent atoms is an absolute minimum. For the purpose of fitting operations made herein, the publicly available software program MOE (Chemical Computing Group Inc.) v. 2002.3 was used.

[00104] The results from this process are typically reported as an RMSD value between two sets of atoms. The term “root mean square deviation” means the square root of the arithmetic mean of the squares of deviations. It is a way to express the deviation or variation

from a trend or object. As used herein, an RMSD value refers to a calculated value based on variations in the atomic coordinates of a reference protein from the atomic coordinates of a reference protein or portions of thereof. The structure coordinates for IspA, provided in Figure 3, are used as the reference protein in these calculations.

**[00105]** The same set of atoms was used for initial fitting of the structures and for computing root mean square deviation values. For example, if a root mean square deviation (RMSD) between C $\alpha$  atoms of two proteins is needed, the proteins in question should be superposed only on the C $\alpha$  atoms and not on any other set of atoms. Similarly, if an RMSD calculation for all atoms is required, the superposition of two structures should be performed on all atoms.

**[00106]** Based on a review of protein structures deposited in the Protein Databank (PDB), 1FPS was identified as having the smallest RMSD values relative to the structure coordinates provided herein. Table 10 below provides a series of RMSD values that were calculated by the above described process using the structure coordinates in Figure 3 as the reference protein and the structure coordinates from PDB code: 1FPS as the target protein.

TABLE 10

AA RESIDUES USED TO PERFORM RMSD COMPARISON WITH PDB: 1FPS	PORTION OF EACH AA RESIDUE USED TO PERFORM RMSD COMPARISON WITH PDB:1FPS	RMSD [Å]
<b>IPP BINDING POCKET</b>		
Table 2 (4 Angstrom set)	alpha-carbon atoms <sup>1</sup>	2.41
	main-chain atoms <sup>1</sup>	2.37
	all non-hydrogen <sup>2</sup>	4.17
Table 3 (7 Angstrom set)	alpha-carbon atoms <sup>1</sup>	3.41
	main-chain atoms <sup>1</sup>	3.47
	all non-hydrogen <sup>2</sup>	4.68
Table 4 (10 Angstrom set)	alpha-carbon atoms <sup>1</sup>	3.02
	main-chain atoms <sup>1</sup>	3.05
	all non-hydrogen <sup>2</sup>	3.87
<b>RISEDRONATE BINDING POCKET</b>		
Table 5 (4 Angstrom set)	alpha-carbon atoms <sup>1</sup>	1.88
	main-chain atoms <sup>1</sup>	1.83
	all non-hydrogen <sup>2</sup>	2.19
Table 6 (7 Angstrom set)	alpha-carbon atoms <sup>1</sup>	2.04
	main-chain atoms <sup>1</sup>	1.99
	all non-hydrogen <sup>2</sup>	2.41
Table 7 (10 Angstrom set)	alpha-carbon atoms <sup>1</sup>	2.66
	main-chain atoms <sup>1</sup>	2.73
	all non-hydrogen <sup>2</sup>	3.61
<b>ENTIRE PROTEIN</b>		
16-314 of SEQ. ID No. 1	alpha-carbon atoms <sup>1</sup>	3.64
	main-chain atoms <sup>1</sup>	3.59
	all non-hydrogen <sup>2</sup>	4.06

<sup>1</sup>- the RMSD computed between the atoms of all amino acids that are common to both the target and the reference in the aligned and superposed structure. The amino acids need not to be identical.

<sup>2</sup>- the RMSD computed only between identical amino acids, which are common to both the target and the reference in the aligned and superposed structure.

[00107] It is noted that mutants, homologs and variants of IspA are likely to have similar structures despite having different sequences. For example, the binding pockets of these related proteins are likely to have similar contours. Accordingly, it should be recognized that the structure coordinates and binding pocket models provided herein have utility for these other related proteins.

[00108] Accordingly, in one embodiment, the invention relates to data, computer readable media comprising data, and uses of the data where the data comprises all or a portion of the structure coordinates shown in Figure 3 or structure coordinates having a root mean square deviation (RMSD) equal to or less than the RMSD value specified in Columns 3, 4 or 5 of Table 1 when compared to the structure coordinates of Figure 3, the root mean square deviation being calculated such that the portion of amino acid residues specified in Column 2 of Table 1 of each set of structure coordinates are superimposed and the root mean square deviation is based only on those amino acid residues in the structure coordinates that are also present in the portion of the protein specified in specified in Column 1 of Table 1.

[00109] As noted, there are many different ways to express the surface contours of the IspA structure other than by using the structure coordinates provided in Figure 3. Accordingly, it is noted that the present invention is also directed to any data, computer readable media comprising data, and uses of the data where the data defines a computer model for a protein binding pocket, at least a portion of the computer model having a surface contour that has a root mean square deviation equal to or less than a given RMSD value specified in Columns 3, 4 or 5 of Table 1 when the coordinates used to compute the surface contour are compared to the structure coordinates of Figure 3, wherein (a) the root mean square deviation is calculated by the calculation method set forth herein, (b) the portion of amino acid residues associated with the given RMSD value in Table 1 (specified in Column 2 of Table 1) are superimposed according to the RMSD calculation, and (c) the root mean square deviation is calculated based only on those amino acid residues present in both the protein being modeled and the portion of the protein associated with the given RMSD in Table 1 (specified in Column 1 of Table 1).

## 5. IspA Structure

[00110] The present invention is also directed to a three-dimensional crystal structure of IspA. This crystal structure may be used to identify binding sites, to provide mutants having

desirable binding properties, and ultimately, to design, characterize, or identify ligands that interact with IspA as well as homologs and other closely related proteins.

[00111] The three-dimensional crystal structure of IspA may be generated, as is known in the art, from the structure coordinates shown in Figure 3 and similar such coordinates.

[00112] During the course of structure solution it became evident that the crystals of IspA of the present invention contained two IspA molecules in the asymmetric unit. These two molecules interact to form the biological dimer.

[00113] The final refined coordinates include amino acid residues 16-314 (Figure 3).

[00114] Figure 4 illustrates a ribbon diagram overview of the structure of IspA, highlighting the secondary structural elements of the protein. Aside from significant differences in ligand and inhibitor-binding active site residues, the overall secondary structure of IspA\_Ec resembles that of avian farnesyl pyrophosphate synthase (pdb code 1FPS; Biochemistry (1994): 33, 10871) which shares 26% sequence identity. The enzyme forms a long and somewhat flat all- $\alpha$  helical structure that packs to form 3 distinct layers. The first layer is formed by helices 1 and 2 and is orthogonal to the two others. The second layer contains helices 3,4,5 and 10, and the third is formed by helices 6, 7, 8, and 9. Helices 4 and 8 are located in the center of the protein core and contain the conserved Asp rich motifs that identify IspA\_Ec as a member of the isoprenoid diphosphate synthase family.

## 6. IspA active site and Ligand Interaction

[00115] The term "binding site" or "binding pocket", as the terms are used herein, refers to a region of a protein that, as a result of its shape, favorably associates with a ligand or substrate. The term "IspA-like binding pocket" refers to a portion of a molecule or molecular complex whose shape is sufficiently similar to the IspA binding pockets as to bind common ligands. This commonality of shape may be quantitatively defined based on a comparison to a reference point, that reference point being the structure coordinates provided herein. For example the commonality of shape may be quantitatively defined based on a root mean square deviation (RMSD) from the structure coordinates of the backbone atoms of the amino acids that make up the binding pockets in IspA (as set forth in Figure 3).

[00116] The "active site binding pockets" or "active site" of IspA refers to the area on the surface of IspA where the substrate binds.

[00117] A major feature of the IspA structure is a large centrally located cavity that binds the allylic and homoallylic isoprenoid substrates as well as bisphosphonate inhibitors. DMAPP and inhibitor binding is mediated by a trinuclear Mg center that is ligated by conserved protein side chains that emanate from the two Aps rich motifs located on opposite walls of the cavity. In the active site all three Mg atoms are octahedrally coordinated by protein, pyrophosphate and water oxygens. The Mg1 site is coordinated by the side chain of Asp244 from the second DDXXD motif, two diphosphate oxygen atoms, and three water molecules. The side chain of Asp105 and Asp111 form the first DDXXD motif, two diphosphate oxygens, and two water molecules coordinate Mg2. Notably, both of these metal centers form 6 membered-ring chelate structures with the risedronate diphosphate. These interactions anchor the inhibitor to the enzyme active site and constrain the geometry of its diphosphate such that the nonbridging oxygens on adjacent phosphates are maintained in a coplanar arrangement. Asp105 and Asp111, as well as a diphosphate oxygen and three water molecules coordinate the third active site Mg (Mg3). In addition to multiple metal coordination interactions, the inhibitor diphosphate makes ionic interactions with the side chains of conserved Arg116, Lys202, and Lys258.

[00118] This risedronate:IPP ternary complex reveals that the FPPS C-terminus participates in catalysis by organizing conserved residues that interact with the IPP pyrophosphate. Notably, a salt bridge between the C-terminus and Lys66 positions its side chain NZ atom in an optimal location to interact with two nonbridging oxygens on adjacent phosphates. Arg318, which is also positioned by a C-terminal salt bridge, forms a water mediated interaction to a single diphosphate oxygen. Additional interactions, including hydrogen bonds from the Gly65 backbone amide and the side chain of His98, and salt bridges between Arg69 and Arg117, stabilize the enzyme bound IPP diphosphate conformation.

[00119] The pyridyl side chain of the bisphosphonate inhibitor binds in a large hydrophobic pocket that accommodates the growing hydrocarbon tail of the isoprenyl product. Stacking interactions with Gln 179 and Lys 202 on one side and Ser 101 and Leu 102 on the other, as well as a hydrogen bond to conserved Thr 203, position the inhibitor pyridyl group within vanderwalls distance of the C1, C2, C3, and C4 atoms of IPP. The pocket extends to the dimer interface with residues from helix 4 (Tyr100, Ser101, His104, and Met110), helix 6 (Met 175 and Cys176), and helix 5 of the adjacent subunit forming the walls of the pocket.

[00120] Figure 5A illustrates residues within 4.0 Å of the IPP binding pocket for IspA. Figure 5B illustrates residues within 4.0 Å of the Risedronate binding pocket for IspA.

[00121] In resolving the crystal structure of IspA, applicants determined that IspA amino acids shown in Table 2 (above) are encompassed within a 4-Angstrom radius around the IspA active site and therefore are likely close enough to interact with an active site inhibitor of IspA. Applicants have also determined that the amino acids shown in Table 3 (above) are encompassed within a 7-Angstrom radius around the IspA active site. Further, the amino acids shown in Table 4 (above) are encompassed within a 10-Angstrom radius around the IspA active site. Due to their proximity to the active site, the amino acids in the 4, 7, and/or 10 Å sets are preferably conserved in variants of IspA. While it is desirable to largely conserve these residues, it should be recognized however that variants may also involve varying 1, 2, 3, 4 or more of the residues set forth in Tables 2, 3, 4, 5, 6 and 7 in order to evaluate the roles these amino acids play in the binding pocket.

[00122] With the knowledge of the IspA crystal structure provided herein, Applicants are able to know the contour of an IspA binding pocket based on the relative positioning of the 4, 7, and/or 10 Å sets of amino acids. Again, it is noted that it may be desirable to form variants where 1, 2, 3, 4 or more of the residues set forth in Tables 2, 3, 4, 5, 6 and 7 are varied in order to evaluate the roles these amino acids play in the binding pocket. Accordingly, any set of structure coordinates for a protein from any source shall be considered within the scope of the present invention if the structure coordinates have a root mean square deviation equal to or less than the RMSD value specified in Columns 3, 4 or 5 of Table 1 when compared to the structure coordinates of Figure 3, the root mean square deviation being calculated such that the portion of amino acid residues specified in Column 2 of Table 1 of each set of structure coordinates are superimposed and the root mean square deviation is based only on those amino acid residues in the structure coordinates that are also present in the portion of the protein specified in specified in Column 1 of Table 1.

[00123] Accordingly, in various embodiments, the invention relates to data, computer readable media comprising data, and uses of the data where the data comprises structure coordinates that have a root mean square deviation equal to or less than the RMSD value specified in Columns 3, 4 or 5 of Table 1 when compared to the structure coordinates of Figure 3, the root mean square deviation being calculated such that the portion of amino acid residues



specified in Column 2 of Table 1 of each set of structure coordinates are superimposed and the root mean square deviation is based only on those amino acid residues in the structure coordinates that are also present in the portion of the protein specified in specified in Column 1 of Table 1.

[00124] As noted above, there are many different ways to express the surface contours of the IspA structure other than by using the structure coordinates provided in Figure 3. Accordingly, it is noted that the present invention is also directed to any data, computer readable media comprising data, and uses of the data where the data defines a computer model for a protein binding pocket, at least a portion of the computer model having a surface contour that has a root mean square deviation equal to or less than a given RMSD value specified in Columns 3, 4 or 5 of Table 1 when the coordinates used to compute the surface contour are compared to the structure coordinates of Figure 3, wherein (a) the root mean square deviation is calculated by the calculation method set forth herein, (b) the portion of amino acid residues associated with the given RMSD value in Table 1 (specified in Column 2 of Table 1) are superimposed according to the RMSD calculation, and (c) the root mean square deviation is calculated based only on those amino acid residues present in both the protein being modeled and the portion of the protein associated with the given RMSD in Table 1 (specified in Column 1 of Table 1).

[00125] It will be readily apparent to those of skill in the art that the numbering of amino acids in other isoforms of IspA may be different than that set forth for IspA. Corresponding amino acids in other isoforms of IspA are easily identified by visual inspection of the amino acid sequences or by using commercially available homology software programs, as further described below.

#### 7. System For Displaying the Three Dimensional Structure of IspA

[00126] The present invention is also directed to machine-readable data storage media having data storage material encoded with machine-readable data that comprises structure coordinates for IspA. The present invention is also directed to a machine readable data storage media having data storage material encoded with machine readable data, which, when read by an appropriate machine, can display a three dimensional representation of a structure of IspA.

[00127] All or a portion of the IspA coordinate data shown in Figure 3, when used in conjunction with a computer programmed with software to translate those coordinates into the three-dimensional structure of IspA may be used for a variety of purposes, especially for purposes relating to drug discovery. Softwares for generating three-dimensional graphical representations are known and commercially available. The ready use of the coordinate data requires that it be stored in a computer-readable format. Thus, in accordance with the present invention, data capable of being displayed as the three-dimensional structure of IspA and/or portions thereof and/or their structurally similar variants may be stored in a machine-readable storage medium, which is capable of displaying a graphical three-dimensional representation of the structure.

[00128] For example, in various embodiments, a computer is provided for producing a three-dimensional representation of at least an IspA-like binding pocket, the computer comprising:

machine readable data storage medium comprising a data storage material encoded with machine-readable data, the machine readable data comprising structure coordinates that have a root mean square deviation equal to or less than the RMSD value specified in Columns 3, 4 or 5 of Table 1 when compared to the structure coordinates of Figure 3, the root mean square deviation being calculated such that the portion of amino acid residues specified in Column 2 of Table 1 of each set of structure coordinates are superimposed and the root mean square deviation is based only on those amino acid residues in the structure coordinates that are also present in the portion of the protein specified in specified in Column 1 of Table 1;

a working memory for storing instructions for processing the machine-readable data;

a central-processing unit coupled to the working memory and to the machine-readable data storage medium, for processing the machine-readable data into the three-dimensional representation; and

an output hardware coupled to the central processing unit, for receiving the three dimensional representation.

[00129] Another embodiment of this invention provides a machine-readable data storage medium, comprising a data storage material encoded with machine readable data which, when used by a machine programmed with instructions for using said data, displays a graphical three-dimensional representation comprising IspA or a portion or variant thereof.

**[00130]** In various variations, the machine readable data comprises data for representing a protein based on structure coordinates where the structure coordinates have a root mean square deviation equal to or less than the RMSD value specified in Columns 3, 4 or 5 of Table 1 when compared to the structure coordinates of Figure 3, the root mean square deviation being calculated such that the portion of amino acid residues specified in Column 2 of Table 1 of each set of structure coordinates are superimposed and the root mean square deviation is based only on those amino acid residues in the structure coordinates that are also present in the portion of the protein specified in specified in Column 1 of Table 1.

**[00131]** According to another embodiment, the machine-readable data storage medium comprises a data storage material encoded with a first set of machine readable data which comprises the Fourier transform of structure coordinates that have a root mean square deviation equal to or less than the RMSD value specified in Columns 3, 4 or 5 of Table 1 when compared to the structure coordinates of Figure 3, the root mean square deviation being calculated such that the portion of amino acid residues specified in Column 2 of Table 1 of each set of structure coordinates are superimposed and the root mean square deviation is based only on those amino acid residues in the structure coordinates that are also present in the portion of the protein specified in specified in Column 1 of Table 1, and which, when using a machine programmed with instructions for using said data, can be combined with a second set of machine readable data comprising the X-ray diffraction pattern of another molecule or molecular complex to determine at least a portion of the structure coordinates corresponding to the second set of machine readable data. For example, the Fourier transform of the structure coordinates set forth in Figure 3 may be used to determine at least a portion of the structure coordinates of other IspA-like enzymes, and isoforms of IspA.

**[00132]** Optionally, a computer system is provided in combination with the machine-readable data storage medium provided herein. In one embodiment, the computer system comprises a working memory for storing instructions for processing the machine-readable data; a processing unit coupled to the working memory and to the machine-readable data storage medium, for processing the machine-readable data into the three-dimensional representation; and an output hardware coupled to the processing unit, for receiving the three-dimensional representation.

[00133] Figure 6 illustrates an example of a computer system that may be used in combination with storage media according to the present invention. As illustrated, the computer system 10 includes a computer 11 comprising a central processing unit ("CPU") 20, a working memory 22 which may be, e.g., RAM (random-access memory) or "core" memory, mass storage memory 24 (such as one or more disk drives or CD-ROM drives), one or more cathode-ray tube ("CRT") display terminals 26, one or more keyboards 28, one or more input lines 30, and one or more output lines 40, all of which are interconnected by a conventional bi-directional system bus 50.

[00134] Input hardware 36, coupled to computer 11 by input lines 30, may be implemented in a variety of ways. For example, machine-readable data of this invention may be inputted via the use of a modem or modems 32 connected by a telephone line or dedicated data line 34. Alternatively or additionally, the input hardware 36 may comprise CD-ROM drives or disk drives 24. In conjunction with display terminal 26, keyboard 28 may also be used as an input device.

[00135] Conventional devices may, similarly implement output hardware 46, coupled to computer 11 by output lines 40. By way of example, output hardware 46 may include CRT display terminal 26 for displaying a graphical representation of a binding pocket of this invention using a program such as MOE as described herein. Output hardware might also include a printer 42, so that hard copy output may be produced, or a disk drive 24, to store system output for later use.

[00136] In operation, CPU 20 coordinates the use of the various input and output devices 36, 46 coordinates data accesses from mass storage 24 and accesses to and from working memory 22, and determines the sequence of data processing steps. A number of programs may be used to process the machine-readable data of this invention. Such programs are discussed in reference to using the three dimensional structure of IspA described herein.

[00137] The storage medium encoded with machine-readable data according to the present invention can be any conventional data storage device known in the art. For example, the storage medium can be a conventional floppy diskette or hard disk. The storage medium can also be an optically readable data storage medium, such as a CD-ROM or a DVD-ROM, or a rewritable medium such as a magneto-optical disk that is optically readable and magneto-optically writable.

## 8. Uses of the Three Dimensional Structure of IspA

[00138] The three-dimensional crystal structure of the present invention may be used to identify IspA binding sites, be used as a molecular replacement model to solve the structure of unknown crystallized proteins, to design mutants having desirable binding properties, and ultimately, to design, characterize, identify entities capable of interacting with IspA and other related homologs, as well as other uses that would be recognized by one of ordinary skill in the art. Such entities may be chemical entities or proteins. The term "chemical entity", as used herein, refers to chemical compounds, complexes of at least two chemical compounds, and fragments of such compounds.

[00139] The IspA structure coordinates provided herein are useful for screening and identifying drugs that inhibit IspA and other related homologs. For example, the structure encoded by the data may be computationally evaluated for its ability to associate with putative substrates or ligands. Such compounds that associate with IspA may inhibit IspA, and are potential drug candidates. Additionally or alternatively, the structure encoded by the data may be displayed in a graphical three-dimensional representation on a computer screen. This allows visual inspection of the structure, as well as visual inspection of the structure's association with the compounds.

[00140] Thus, according to another embodiment of the present invention, a method is provided for evaluating the potential of an entity to associate with IspA or a fragment or variant thereof by using all or a portion of the structure coordinates provided in Figure 3 or functional equivalents thereof. A method is also provided for evaluating the potential of an entity to associate with IspA or a fragment or variant thereof by using structure coordinates similar to all or a portion of the structure coordinates provided in Figure 3 or functional equivalents thereof.

[00141] The method may optionally comprise the steps of: creating a computer model of all or a portion of a protein structure (e.g., a binding pocket) using structure coordinates according to the present invention; performing a fitting operation between the entity and the computer model; and analyzing the results of the fitting operation to quantify the association between the entity and the model. The portion of the protein structure used optionally comprises all of the amino acids listed in Tables 2, 3, 4, 5, 6 and 7 that are present in the structure coordinates being used.

[00142] It is noted that the computer model may not necessarily directly use the structure coordinates. Rather, a computer model can be formed that defines a surface contour that is the same or similar to the surface contour defined by the structure coordinates.

[00143] The structure coordinates provided herein can also be utilized in a method for identifying a ligand (e.g., entities capable of associating with a protein) of a protein comprising an IspA-like binding pocket. One embodiment of the method comprises: using all or a portion of the structure coordinates provided herein to generate a three-dimensional structure of an IspA-like binding pocket; employing the three-dimensional structure to design or select a potential ligand; synthesizing the potential ligand; and contacting the synthesized potential ligand with a protein comprising an IspA-like binding pocket to determine the ability of the potential ligand to interact with protein. According to this method, the structure coordinates used may have a root mean square deviation equal to or less than the RMSD values specified in Columns 3, 4 or 5 of Table 1 when compared to the structure coordinates of Figure 3 according to the RMSD calculation method set forth herein, provided that the portion of amino acid residues specified in Column 2 of Table 1 of each set of structure coordinates are superimposed and the root mean square deviation is calculated based only on those amino acid residues in the structure coordinates that are also present in the portion of the protein specified in Column 1 of Table 1. The portion of the protein structure used optionally comprises all of the amino acids listed in Tables 1, 2 and/or 3 that are present.

[00144] As noted previously, the three-dimensional structure of an IspA-like binding pocket need not be generated directly from structure coordinates. Rather, a computer model can be formed that defines a surface contour that is the same or similar to the surface contour defined by the structure coordinates.

[00145] A method is also provided for evaluating the ability of an entity, such as a compound or a protein to associate with an IspA-like binding pocket, the method comprising: constructing a computer model of a binding pocket defined by structure coordinates that have a root mean square deviation equal to or less than the RMSD value specified in Columns 3, 4 or 5 of Table 1 when compared to the structure coordinates of Figure 3, the root mean square deviation being calculated such that the portion of amino acid residues specified in Column 2 of Table 1 of each set of structure coordinates are superimposed and the root mean square deviation is based only on those amino acid residues in the structure coordinates that are also

present in the portion of the protein specified in specified in Column 1 of Table 1; selecting an entity to be evaluated by a method selected from the group consisting of (i) assembling molecular fragments into the entity, (ii) selecting an entity from a small molecule database, (iii) *de novo* ligand design of the entity; and (iv) modifying a known ligand for IspA, or a portion thereof; performing a fitting program operation between computer models of the entity to be evaluated and the binding pocket in order to provide an energy-minimized configuration of the entity in the binding pocket; and evaluating the results of the fitting operation to quantify the association between the entity and the binding pocket model in order to evaluate the ability of the entity to associate with the said binding pocket.

[00146] The computer model of a binding pocket used in this embodiment need not be generated directly from structure coordinates. Rather, a computer model can be formed that defines a surface contour that is the same or similar to the surface contour defined by the structure coordinates.

[00147] Also according to the method, the method may further include synthesizing the entity; and contacting a protein having an IspA-like binding pocket with the synthesized entity.

[00148] With the structure provided herein, the present invention for the first time permits the use of molecular design techniques to identify, select or design potential inhibitors of IspA, based on the structure of an IspA-like binding pocket. Such a predictive model is valuable in light of the high costs associated with the preparation and testing of the many diverse compounds that may possibly bind to the IspA protein.

[00149] According to this invention, a potential IspA inhibitor may now be evaluated for its ability to bind an IspA-like binding pocket prior to its actual synthesis and testing. If a proposed entity is predicted to have insufficient interaction or association with the binding pocket, preparation and testing of the entity can be obviated. However, if the computer modeling indicates a strong interaction, the entity may then be obtained and tested for its ability to bind.

[00150] A potential inhibitor of an IspA-like binding pocket may be computationally evaluated using a series of steps in which chemical entities or fragments are screened and selected for their ability to associate with the IspA-like binding pockets.

[00151] One skilled in the art may use one of several methods to screen entities (whether chemical or protein) for their ability to associate with an IspA-like binding pocket. This

process may begin by visual inspection of, for example, an IspA-like binding pocket on a computer screen based on the IspA structure coordinates in Figure 3 or other coordinates which define a similar shape generated from the machine-readable storage medium. Selected fragments or chemical entities may then be positioned in a variety of orientations, or docked, within that binding pocket as defined above. Docking may be accomplished using software such as Quanta and Sybyl, followed by energy minimization and molecular dynamics with standard molecular mechanics force fields, such as CHARMM and AMBER.

[00152] Specialized computer programs may also assist in the process of selecting entities. These include: GRID (P. J. Goodford, "A Computational Procedure for Determining Energetically Favorable Binding Sites on Biologically Important Macromolecules", J. Med. Chem., 28, pp. 849-857 (1985)). GRID is available from Oxford University, Oxford, UK; MCSS (A. Miranker et al., "Functionality Maps of Binding Sites: A Multiple Copy Simultaneous Search Method." Proteins: Structure, Function and Genetics, 11, pp. 29-34 (1991)). MCSS is available from Molecular Simulations, San Diego, Calif.; AUTODOCK (D. S. Goodsell et al., "Automated Docking of Substrates to Proteins by Simulated Annealing", Proteins: Structure, Function, and Genetics, 8, pp. 195-202 (1990)). AUTODOCK is available from Scripps Research Institute, La Jolla, Calif.; & DOCK (I. D. Kuntz et al., "A Geometric Approach to Macromolecule-Ligand Interactions", J. Mol. Biol., 161, pp. 269-288 (1982)). DOCK is available from University of California, San Francisco, Calif.

[00153] Once suitable entities have been selected, they can be designed or assembled. Assembly may be preceded by visual inspection of the relationship of the fragments to each other on the three-dimensional image displayed on a computer screen in relation to the structure coordinates of IspA. This may then be followed by manual model building using software such as MOE, QUANTA or Sybyl [Tripos Associates, St. Louis, Mo].

[00154] Useful programs to aid one of skill in the art in connecting the individual chemical entities or fragments include: CAVEAT (P. A. Bartlett et al, "CAVEAT: A Program to Facilitate the Structure-Derived Design of Biologically Active Molecules", in "Molecular Recognition in Chemical and Biological Problems", Special Pub., Royal Chem. Soc., 78, pp. 182-196 (1989); G. Lauri and P. A. Bartlett, "CAVEAT: a Program to Facilitate the Design of Organic Molecules", J. Comput. Aided Mol. Des., 8, pp. 51-66 (1994)). CAVEAT is available from the University of California, Berkeley, Calif.; 3D Database systems such as ISIS (MDL



Information Systems, San Leandro, Calif.). This area is reviewed in Y. C. Martin, "3D Database Searching in Drug Design", J. Med. Chem., 35, pp. 2145-2154 (1992); HOOK (M. B. Eisen et al, "HOOK: A Program for Finding Novel Molecular Architectures that Satisfy the Chemical and Steric Requirements of a Macromolecule Binding Site", Proteins: Struct., Funct., Genet., 19, pp. 199-221 (1994). HOOK is available from Molecular Simulations, San Diego, Calif.

[00155] Instead of proceeding to build an inhibitor of an IspA-like binding pocket in a step-wise fashion one fragment or entity at a time as described above, inhibitory or other IspA binding compounds may be designed as a whole or "*de novo*" using either an empty binding site or optionally including some portion(s) of a known inhibitor(s). There are many *de novo* ligand design methods including: LUDI (H.-J. Bohm, "The Computer Program LUDI: A New Method for the De Novo Design of Enzyme Inhibitors", J. Comp. Aid. Molec. Design, 6, pp. 61-78 (1992)). LUDI is available from Molecular Simulations Incorporated, San Diego, Calif.; LEGEND (Y. Nishibata et al., Tetrahedron, 47, p. 8985 (1991)). LEGEND is available from Molecular Simulations Incorporated, San Diego, Calif.; LEAPFROG (available from Tripos Associates, St. Louis, Mo.); & SPROUT (V. Gillet et al, "SPROUT: A Program for Structure Generation)", J. Comput. Aided Mol. Design, 7, pp. 127-153 (1993)). SPROUT is available from the University of Leeds, UK.

[00156] Other molecular modeling techniques may also be employed in accordance with this invention (see, e.g., Cohen et al., "Molecular Modeling Software and Methods for Medicinal Chemistry, J. Med. Chem., 33, pp. 883-894 (1990); see also, M. A. Navia and M. A. Murcko, "The Use of Structural Information in Drug Design", Current Opinions in Structural Biology, 2, pp. 202-210 (1992); L. M. Balbes et al., "A Perspective of Modern Methods in Computer-Aided Drug Design", in Reviews in Computational Chemistry, Vol. 5, K. B. Lipkowitz and D. B. Boyd, Eds., VCH, New York, pp. 337-380 (1994); see also, W. C. Guida, "Software For Structure-Based Drug Design", Curr. Opin. Struct. Biology, 4, pp. 777-781 (1994)).

[00157] Once an entity has been designed or selected, for example, by the above methods, the efficiency with which that entity may bind to an IspA binding pocket may be tested and optimized by computational evaluation. For example, an effective IspA binding pocket inhibitor preferably demonstrates a relatively small difference in energy between its

bound and free states (i.e., a small deformation energy of binding). Thus, the most efficient IspA binding pocket inhibitors should preferably be designed with deformation energy of binding of not greater than about 10 kcal/mole, more preferably, not greater than 7 kcal/mole. IspA binding pocket inhibitors may interact with the binding pocket in more than one of multiple conformations that are similar in overall binding energy. In those cases, the deformation energy of binding is taken to be the difference between the energy of the free entity and the average energy of the conformations observed when the inhibitor binds to the protein.

[00158] An entity designed or selected as binding to an IspA binding pocket may be further computationally optimized so that in its bound state it would preferably lack repulsive electrostatic interaction with the target enzyme and with the surrounding water molecules. Such non-complementary electrostatic interactions include repulsive charge-charge, dipole-dipole and charge-dipole interactions.

[00159] Specific computer software is available in the art to evaluate compound deformation energy and electrostatic interactions. Examples of programs designed for such uses include: Gaussian 94, revision C (M. J. Frisch, Gaussian, Inc., Pittsburgh, Pa. COPYRIGHT.1995); AMBER, version 4.1 (P. A. Kollman, University of California at San Francisco, COPYRIGHT 1995); QUANTA/CHARMM (Molecular Simulations, Inc., San Diego, Calif. COPYRIGHT.1995); Insight II/Discover (Molecular Simulations, Inc., San Diego, Calif. COPYRIGHT.1995); DelPhi (Molecular Simulations, Inc., San Diego, Calif. COPYRIGHT.1995); and AMSOL (Quantum Chemistry Program Exchange, Indiana University). These programs may be implemented, for instance, using a Silicon Graphics workstation such as an Indigo.sup.2 with "IMPACT" graphics. Other hardware systems and software packages will be known to those skilled in the art.

[00160] Another approach provided by this invention, is the computational screening of small molecule databases for chemical entities or compounds that can bind in whole, or in part, to an IspA binding pocket. In this screening, the quality of fit of such entities to the binding site may be judged either by shape complementarities or by estimated interaction energy [E. C. Meng et al., J. Comp. Chem., 13, 505-524 (1992)].

[00161] According to another embodiment, the invention provides compounds that associate with an IspA-like binding pocket produced or identified by various methods set forth above.

[00162] The structure coordinates set forth in Figure 3 can also be used to aid in obtaining structural information about another crystallized molecule or molecular complex. This may be achieved by any of a number of well-known techniques, including molecular replacement.

[00163] For example, a method is also provided for utilizing molecular replacement to obtain structural information about a protein whose structure is unknown comprising the steps of: generating an X-ray diffraction pattern of a crystal of the protein whose structure is unknown; generating a three-dimensional electron density map of the protein whose structure is unknown from the X-ray diffraction pattern by using at least a portion of the structure coordinates set forth in Figure 3 as a molecular replacement model.

[00164] By using molecular replacement, all or part of the structure coordinates of the IspA provided by this invention (and set forth in Figure 3) can be used to determine the structure of another crystallized molecule or molecular complex more quickly and efficiently than attempting an *ab initio* structure determination. One particular use includes use with other related IspA homologs. Molecular replacement provides an accurate estimation of the phases for an unknown structure. Phases are a factor in equations used to solve crystal structures that cannot be determined directly. Obtaining accurate values for the phases, by methods other than molecular replacement, is a time-consuming process that involves iterative cycles of approximations and refinements and greatly hinders the solution of crystal structures. However, when the crystal structure of a protein containing at least a homologous portion has been solved, the phases from the known structure provide a satisfactory estimate of the phases for the unknown structure.

[00165] Thus, this method involves generating a preliminary model of a molecule or molecular complex whose structure coordinates are unknown, by orienting and positioning the relevant portion of IspA according to Figure 3 within the unit cell of the crystal of the unknown molecule or molecular complex so as best to account for the observed X-ray diffraction pattern of the crystal of the molecule or molecular complex whose structure is unknown. Phases can then be calculated from this model and combined with the observed X-ray diffraction pattern

amplitudes to generate an electron density map of the structure whose coordinates are unknown. This, in turn, can be subjected to any well-known model building and structure refinement techniques to provide a final, accurate structure of the unknown crystallized molecule or molecular complex [E. Lattman, "Use of the Rotation and Translation Functions", in *Meth. Enzymol.*, 115, pp. 55-77 (1985); M. G. Rossmann, ed., "The Molecular Replacement Method", *Int. Sci. Rev. Ser.*, No. 13, Gordon & Breach, New York (1972)].

[00166] The structure of any portion of any crystallized molecule or molecular complex that is sufficiently homologous to any portion of IspA can be resolved by this method.

[00167] In one embodiment, the method of molecular replacement is utilized to obtain structural information about the present invention and any other IspA-like molecule. The structure coordinates of IspA, as provided by this invention, are particularly useful in solving the structure of other isoforms of IspA or IspA complexes.

[00168] The structure coordinates of IspA as provided by this invention are useful in solving the structure of IspA variants that have amino acid substitutions, additions and/or deletions (referred to collectively as "IspA mutants", as compared to naturally occurring IspA). These IspA mutants may optionally be crystallized in co-complex with a ligand, such as an inhibitor, substrate analogue or a suicide substrate. The crystal structures of a series of such complexes may then be solved by molecular replacement and compared with that of IspA. Potential sites for modification within the various binding sites of the enzyme may thus be identified. This information provides an additional tool for determining the most efficient binding interactions such as, for example, increased hydrophobic interactions, between IspA and a ligand. It is noted that the ligand may be the protein's natural ligand or may be a potential agonist or antagonist of a protein.

[00169] All of the complexes referred to above may be studied using well-known X-ray diffraction techniques and may be refined versus 1.5-3Å resolution X-ray data to an R value of about 0.22 or less using computer software, such as X-PLOR [Yale University, COPYRIGHT.1992, distributed by Molecular Simulations, Inc.; see, e.g., Blundell & Johnson, *supra*; *Meth. Enzymol.*, Vol. 114 & 115, H. W. Wyckoff et al., eds., Academic Press (1985)]. This information may thus be used to optimize known IspA inhibitors, and more importantly, to design new IspA inhibitors.

[00170] The structure coordinates described above may also be used to derive the dihedral angles, phi and psi, that define the conformation of the amino acids in the protein backbone. As will be understood by those skilled in the art, the  $\phi_n$  angle refers to the rotation around the bond between the alpha-carbon and the nitrogen, and the  $\psi_n$  angle refers to the rotation around the bond between the carbonyl carbon and the alpha-carbon. The subscript "n" identifies the amino acid whose conformation is being described [for a general reference, see Blundell and Johnson, Protein Crystallography, Academic Press, London, 1976].

## 9. Uses of the Crystal and Diffraction Pattern of IspA

[00171] Crystals, crystallization conditions and the diffraction pattern of IspA that can be generated from the crystals also have a range of uses. One particular use relates to screening entities that are not known ligands of IspA for their ability to bind to IspA. For example, with the availability of crystallization conditions, crystals and diffraction patterns of IspA provided according to the present invention, it is possible to take a crystal of IspA; expose the crystal to one or more entities that may be a ligand of IspA; and determine whether a ligand/ IspA complex is formed. The crystals of IspA may be exposed to potential ligands by various methods, including but not limited to, soaking a crystal in a solution of one or more potential ligands or co-crystallizing IspA in the presence of one or more potential ligands. Given the structure coordinates provided herein, once a ligand complex is formed, the structure coordinates can be used as a model in molecular replacement in order to determine the structure of the ligand complex.

[00172] Once one or more ligands are identified, structural information from the ligand/ IspA complex(es) may be used to design new ligands that bind tighter, bind more specifically, have better biological activity or have better safety profile than known ligands.

[00173] In one embodiment, a method is provided for identifying a ligand that binds to IspA comprising: (a) attempting to crystallize a protein that comprises a sequence with 55%, 65%, 78%, 85%, 90%, 95%, 97%, 99% or greater identity with SEQ. ID No. 1 in the presence of one or more entities; (b) if crystals of the protein are obtained in step (a), obtaining an X-ray diffraction pattern of the protein crystal; and (c) determining whether a ligand/protein complex was formed by comparing an X-ray diffraction pattern of a crystal of the protein formed in the

absence of the one or more entities to the crystal formed in the presence of the one or more entities.

[00174] In another embodiment, a method is provided for identifying a ligand that binds to IspA comprising: soaking a crystal of a protein that comprises a sequence with 55%, 65%, 78%, 85%, 90%, 95%, 97%, 99% or greater identity with SEQ. ID No. 1 with one or more entities; determining whether a ligand/protein complex was formed by comparing an X-ray diffraction pattern of a crystal of the protein that has not been soaked with the one or more entities to the crystal that has been soaked with the one or more entities.

[00175] Optionally, the method may further comprise converting the diffraction patterns into electron density maps using phases of the protein crystal and comparing the electron density maps.

[00176] Libraries of "shape-diverse" compounds may optionally be used to allow direct identification of the ligand-receptor complex even when the ligand is exposed as part of a mixture. According to this variation, the need for time-consuming de-convolution of a hit from the mixture is avoided. More specifically, the calculated electron density function reveals the binding event, identifies the bound compound and provides a detailed 3-D structure of the ligand-receptor complex. Once a hit is found, one may optionally also screen a number of analogs or derivatives of the hit for tighter binding or better biological activity by traditional screening methods. The hit and information about the structure of the target may also be used to develop analogs or derivatives with tighter binding or better biological activity. It is noted that the ligand-IspA complex may optionally be exposed to additional iterations of potential ligands so that two or more hits can be linked together to make a more potent ligand. Screening for potential ligands by co-crystallization and/or soaking is further described in U.S. Patent No. 6,297,021, which is incorporated herein by reference.

## EXAMPLES

### **Example 1. Expression and Purification of IspA<sub>Ec</sub>**

[00177] This example describes the expression of IspA<sub>Ec</sub>. It should be noted that a variety of other expression systems and hosts are also suitable for the expression of IspA<sub>Ec</sub>, as would be readily appreciated by one of skill in the art.

[00178] The gene encoding residues 1-299 (from SEQ. ID No. 1), which corresponds to the full-length IspA from E.coli, was isolated by PCR from E. coli genomic DNA (DH10B-1r strain) and cloned into the TOPO-activated cloning site of pSX28 vector. This DNA sequence is presented in SEQ. ID No. 2. Expression in this vector generated a fusion of the full-length IspA with non-cleavable amino-terminal six histidine tag, the amino acid sequence of which is shown, underlined, in Figure 1 (SEQ. ID No. 3).

[00179] Biomass for purification of recombinant IspA\_Ec was generated using 96-well fermentor. Cells from a single 70 ml fermentor tubes was thawed by addition of 21 ml of lysis buffer (50 mM Tris/HCl pH 7.9, 50 mM NaCl, 1 mM MgCl<sub>2</sub>) containing hen egg white lysozyme (0.6 mg/ml) and Benzonase (2.5 U/ml) and sonicated using Sonic Hedgehog robot. The sonicate was allowed to stand for 30 minutes at ~4°C. Total lysate was clarified by centrifugation and 2mL of 5M NaCl were added to the cleared lysate. The cleared lysate from a single fermentor tubes was applied to 3 ml bed ProBond column that had been equilibrated to 50 mM Potassium Phosphate pH 7.8, 0.4 M NaCl, 0.1 M KCl, 20 mM imidazole, 10% glycerol, 0.25 mM TCEP. The solution was passed through the column using gravity flow and the column was washed with 6 bed volumes of 50 mM Potassium Phosphate pH 7.8, 0.4 M NaCl, 0.1 M KCl, 40 mM imidazole, 10% glycerol, 0.25 mM TCEP. The product was eluted with 12 ml of 50 mM Potassium Phosphate pH 7.4, 0.4 M NaCl, 0.1 M KCl, 200 mM imidazole, 10% glycerol, 0.25mM TCEP. The eluted protein was concentrated and buffer-exchanged into 25 mM Tris pH 7.9, 150 mM NaCl by using Vivaspinn centrifugal concentrators. Following three five-fold dilution buffer-exchanges, the IMAC purified IspA\_Ec was concentrated to 12.1 mg/ml with a total volume of 1.08 ml. The purified protein had the correct molecular mass as determined by Mass Spectrograph (MS) analysis (33,812 observed and 33,810 expected without N-terminal methionine), was monomeric by analytical size-exclusion chromatography (SEC) and exhibited a major band by both isoelectric focusing (IEF) and by sodium-dodecyl-sulfate polyacrylamide gel electrophoresis (SDS-PAGE) analyses.

[00180] The portion of the gene encoding residues 544-935 (from SEQ. ID No. 1), which corresponds to the catalytic domain of human IspA, was cloned into a modified pFastBacHTc vector (also known as pSXB1) at the *Bam*HI and *Xba*I sites. The region corresponding to amino acid residues 694-753 (SEQ. ID No. 1) or nucleotide sequence 451-630 (SEQ. ID No. 1) was deleted by using inverse PCR, which generated additional Thr and Ser residues at positions

182-183, respectively (SEQ. ID No. 2). Expression from this vector produced the recombinant IspA catalytic domain with a 6x-histidine tag at the N-terminus followed by a rTEV protease cleavage sequence to facilitate tag removal (the excised 6x-Histidine tag and rTev cleavage site sequences are underlined in SEQ. ID No. 2). Recombinant baculovirus genomic DNAs incorporating the IspA catalytic domain cDNA sequences were generated by transposition using the Bac-to-Bac system (Invitrogen). Infectious viral particles were obtained by transfection of a 2 ml adherent culture of *Spodoptera frugiperda* Sf9 insect cells with the recombinant viral genomic DNA. Growth in ESF 921 protein free medium (Expression Systems) was for 3 days at 27°C. The resulting Passage 0 viral supernatant was used to obtain Passage 1 high titer viral stock (HTS) by infection of a 30 ml adherent culture of *Spodoptera frugiperda* Sf9 insect cells grown under similar conditions. Passage 1 HTS was used in turn to infect a 100 ml suspension culture of *Spodoptera frugiperda* Sf9 insect cells in order to generate Passage 2 HTS.

[00181] Passage 2 HTS was used to infect a 5-liter culture of *Spodoptera frugiperda* Sf9 insect cells (at a density of approx.  $3 \times 10^6$  cells/ml) in a 10 liter Wave BioReactor grown in ESF-921 serum-free medium at a multiplicity of infection (moi) of approximately 5 (empirical value based on usual HTS viral counts). Cell growth/infection proceeded for two days after which time the cells were pelleted by centrifugation and the cell pellet stored at -80C until required. Frozen cell pellets from two such 5-liter cultures were removed from the -80C freezer and each suspended in 150 ml of Lysis Buffer (50 mM Tris-HCl, pH 7.9, 200 mM NaCl, 0.25 mM TCEP, 1 mM PMSF and 2 'Complete-EDTA' Roche Protease Inhibitor tablets). The suspensions were stirred for 45 min at 4C followed by centrifugation at 7,000g for 1h. To each supernatant were added 8 ml of a 50% slurry of ProBond (InVitrogen) resin that had been equilibrated in Lysis Buffer without protease inhibitors. The suspensions were mixed for 90 min followed by centrifugation at 640g for 5 min. The supernatants were discarded and the resin pellets washed three times with 50 mM potassium phosphate, pH 7.9, 400 mM NaCl, 0.25 mM TCEP and 1 ug/mL leupeptin. Each resin sample was transferred to an OMNI chromatography column (10 cm x 1.5 cm diameter) at 4C and washed with 50 column volumes of 50 mM potassium phosphate, pH 7.9, 400 mM NaCl, 20 mM imidazole-HCl, pH 7.9, 0.25 mM TCEP and 1 ug/mL leupeptin. The columns were subsequently washed with 5 column volumes of 50 mM Tris-HCl, pH 7.9, 400 mM NaCl, 0.25 mM TCEP and 1



ug/mL leupeptin. Target elution was effected by the addition of 50 mM Tris-HCl, pH 7.9, 400 mM NaCl, 200 mM imidazole-HCl, pH 7.9, 0.25 mM TCEP, 1 ug/mL leupeptin. The eluates were pooled (the yield at this stage was 25.3 mg total protein in 36 ml) and the polyhistidine purification tag removed by cleavage overnight with 100u/ml TEV protease during dialysis against 50 mM Tris-HCl, pH 7.9, 400 mM NaCl, 20 mM imidazole-HCl, pH 7.9, 0.25 mM TCEP and 1 ug/mL leupeptin at 4C. The TEV-treated sample was passed by gravity flow through an 8 ml bed volume of ProBond chelating resin charged with Ni that had been equilibrated in 50 mM Tris-HCl, pH 7.9, 400 mM NaCl, 20 mM imidazole-HCl, pH 7.9, 0.25 mM TCEP and 1 ug/mL leupeptin at 4C. The unbound flow-through material was concentrated and buffer-exchanged into 25 mM Tris-HCl buffer, pH 7.6, 250 mM NaCl, 5 mM DTT and 1 mM EDTA-NaOH, pH 8.0, by using Vivaspin centrifugal concentrators. Following three five-fold dilution buffer-exchanges, the purified IspA was concentrated to 10.6 mg/ml with a total volume of 1.68 ml (17.8 mg purified IspA). The purified protein had the correct molecular mass as determined by Mass Spectrograph (MS) analysis (38,705 expected and 38,700 observed), was monomeric by analytical size-exclusion chromatography (SEC) and exhibited a major band by both isoelectric focusing (IEF) and by sodium-dodecyl-sulfate polyacrylamide gel electrophoresis (SDS-PAGE) analyses.

#### **Example 2. Crystallization of IspA**

[00182] This example describes the crystallization of IspA. It is noted that the precise crystallization conditions used may be further varied, for example by performing a fine screen based on these crystallization conditions.

[00183] IspA protein samples (corresponding to SEQ. ID No. 1) were concentrated to the final concentration of 12mg/ml, incubated with 0.5-2.50 mM MgCl<sub>2</sub> and ligands before initiating crystallization trials. Several combinations of ligands in the 0.5-2.5mM range produced crystals useful for structural analysis. The ligand combinations included: 1) isopentyl pyrophosphate (IPP) + dimethylallyl S-thiolodiphosphate (DMASPP), 2) IPP + farnesyl S-thiolodiphosphate (FSPP), 3) geranyl diphosphate (GPP), 4) IPP + geranyl S-thiolodiphosphate, 5) IPP + Risedronate, 6) IPP + Pamidronate, and 7) Risedronate. Interestingly, it was found that crystallization was facilitated by the presence of ligands.

**[00184]** 50nL of protein was mixed with 50nL of reservoir solution and incubated in sitting drops over a period of one month using the vapor diffusion method. Crystals were obtained after an extensive and broad screen of conditions, followed by optimization. The reservoir conditions which produced crystal used for data collection were: 10% methyl pentanediol (MPD), 0.1M MES pH 6.0

**[00185]** Single crystals were transferred, briefly, into a cryoprotecting solution containing the reservoir solution supplemented with 25% v/v ethylene glycol. Crystals were then flash frozen by immersion in liquid nitrogen and then stored under liquid nitrogen. A crystal of IspA complexed with IPP + Risedronate is illustrated in Figure 2.

**[00186]** While the present invention is disclosed with reference to certain embodiments and examples detailed above, it is to be understood that these embodiments and examples are intended to be illustrative rather than limiting, as it is contemplated that modifications will readily occur to those skilled in the art, which modifications are intended to be within the scope of the invention and the appended claims. All patents, papers, and books cited in this application are incorporated herein in their entirety.